

Bedeutung des spitalinternen  
Entscheidungsprozesses in einem  
Prinzipal-Agent-Modell  
mit doppeltem Unterstellungsverhältnis

**Dissertation**  
**der Wirtschaftswissenschaftlichen Fakultät**  
**der Universität Zürich**

zur Erlangung der Würde  
eines Doktors  
der Wirtschaftswissenschaften  
in Volkswirtschaftslehre

vorgelegt von  
Simon Spika  
von Deutschland

genehmigt im September 2012 auf Antrag von

Prof. Dr. Peter Zweifel  
Prof. Dr. Christian Ernst

Der Lehrbereich Ökonomie der Wirtschaftswissenschaftlichen Fakultät der Universität Zürich gestattet hierdurch die Drucklegung der vorliegenden Dissertation, ohne damit zu den darin ausgesprochenen Anschauungen Stellung zu nehmen.

Zürich, 19. September 2012

Der Vorsitzende des Doktoratsausschusses: Prof. Dr. Dieter Pfaff

# Danksagung

Ein beträchtlicher Teil dieser Arbeit ist im Zug auf der Strecke Konstanz – Zürich, die ich in den letzten fünf Jahren viermal in der Woche von unserer Wohnung in Konstanz zu meinem Arbeitsplatz am Universitätsspital Zürich zurückgelegt habe, entstanden. Ist die Entstehung dieser Arbeit deswegen mit einer Zugfahrt zu vergleichen? Vielleicht nicht ganz, denn bei einer Zugfahrt geht es immer vorwärts und man weiss bereits beim Einsteigen, wo und wann man ankommen wird.

Es gibt eine Reihe von Personen, die sehr viel zur Entstehung der Arbeit beigetragen haben und denen ich hier danken möchte:

Herrn Prof. Dr. Peter Zweifel danke ich sehr herzlich für die Betreuung der Dissertation, die Unterstützung und die vielen Anregungen.

Bei Herrn Prof. Dr. Christian Ernst möchte ich mich sehr dafür bedanken, dass er die Rolle des Koreferenten übernommen hat.

Grosser Dank gilt auch Kim Starzacher und Hugo Keune vom Universitätsspital Zürich, die mir beide durch grosse Flexibilität und durch grosszügiges Entgegenkommen bei meinen Arbeitszeiten sehr geholfen haben.

Ganz herzlich bedanken möchte ich mich auch bei meinen Eltern und Schwiegereltern, die an vielen Wochenenden die Kinder in Schach gehalten und so dazu beigetragen haben, dass ich mich auf die Arbeit konzentrieren konnte.

Mein ganz besonderer Dank gilt meiner Frau Irina. Sie hat mich in den anstrengenden Zeiten während der Entstehung der Arbeit unterstützt und war für mich und die Kinder da. Ausserdem danke ich ihr für das mühsame Korrekturlesen.

Simon Spika

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
1.1	Unvollständige Information auf dem Markt für Spitalleistungen . . . . .	1
1.2	Unzulänglichkeiten der regulativen Eingriffe in die Preisbildung . . . . .	2
1.3	Das Prinzipal-Agent-Modell in der ökonomischen Theorie der Regulierung .	3
<b>2</b>	<b>Fragestellung und Gliederung</b>	<b>4</b>
2.1	Fragestellung . . . . .	4
2.2	Gliederung . . . . .	6
<b>3</b>	<b>Einführung in das Prinzipal-Agent-Modell</b>	<b>8</b>
3.1	Einordnung des Prinzipal-Agent-Modells in die Anreiztheorie . . . . .	8
3.2	Struktur des Prinzipal-Agent-Modells . . . . .	9
3.2.1	Zielfunktionen . . . . .	9
3.2.2	Anreizverträglichkeitsbedingungen und Arten der asymmetrischen Information . . . . .	10
3.2.2.1	Anreizverträglichkeitsbedingung bei versteckter Handlung	10
3.2.2.2	Anreizverträglichkeitsbedingung bei versteckter Information	10
3.2.3	Teilnahmebedingung . . . . .	11
3.2.4	Zeitlicher Ablauf . . . . .	12
3.3	Ergebnis des Prinzipal-Agent-Modells . . . . .	13
<b>4</b>	<b>Divergenzen der Ziele von Spital und Auftraggeber sowie Begründung der Zielfunktion des Spitals</b>	<b>14</b>
4.1	Zielfunktionen und Divergenzen der Ziele von Auftraggeber und Spital im Modell von Chalkley und Malcomson . . . . .	14
4.1.1	Zielfunktion des Auftraggebers . . . . .	14
4.1.2	Zielfunktion des Spitals . . . . .	15
4.1.3	Optimale Anreize und Divergenzen der Ziele . . . . .	15
4.1.3.1	Bedingungen für die first best-Allokation . . . . .	15
4.1.3.2	Anreizverträglichkeitsbedingungen des Spitals . . . . .	16
4.1.3.3	Divergenzen der Ziele von Auftraggeber und Spital . . . .	17
4.2	Begründung der Zielfunktion des Spitals . . . . .	18
4.2.1	Verhalten des Spitals als Ergebnis der Präferenzen mehrerer Ent- scheidungsträger und des spitalinternen Entscheidungsmechanismus	19
4.2.2	Auswirkung des internen Entscheidungsmechanismus am Beispiel der Zielfunktion des Spitals im Modell von Chalkley und Malcomson	21

<b>5</b>	<b>Das Prinzipal-Agent-Modell mit doppeltem Unterstellungsverhältnis</b>	<b>24</b>
5.1	Überblick . . . . .	24
5.2	Spitalinterne Entscheidungsträger Arzt und Management . . . . .	24
5.3	Patienten und Schwere der Erkrankung . . . . .	26
5.3.1	Nachfrage . . . . .	26
5.4	Feststellung der Fallschwere und Behandlung durch den Arzt . . . . .	26
5.5	Behandlungskosten . . . . .	27
5.6	Vertrag zwischen Auftraggeber und Spital sowie Vergütung des Spitals . .	28
5.7	Spitalinterne Fallkostenbudgetierung und frei verfügbare Mittel für den Arzt	29
5.8	Präferenzen und Zielfunktionen . . . . .	30
5.8.1	Qualität der Behandlung . . . . .	30
5.8.2	Präferenzen und Zielfunktion des Auftraggebers . . . . .	31
5.8.3	Präferenzen und Zielfunktion des Arztes . . . . .	34
5.8.4	Präferenzen und Zielfunktion des Managements . . . . .	37
5.8.4.1	Spitalträger . . . . .	37
5.8.4.2	Spitaldirektion . . . . .	38
5.8.4.3	Zielfunktion Management . . . . .	39
5.9	Teilnahmebedingungen . . . . .	40
5.9.1	Teilnahmebedingung des Arztes . . . . .	40
5.9.2	Teilnahmebedingungen des Managements . . . . .	41
5.10	Mechanismen der Entscheidungsfindung . . . . .	42
5.11	Weitere Annahmen . . . . .	44
<b>6</b>	<b>Asymmetrische Information hinsichtlich der Behandlungsqualität</b>	<b>45</b>
6.1	Praktischer Hintergrund und Abhandlung in der Theorie . . . . .	45
6.1.1	Problematische Messbarkeit der Behandlungsqualität in der Praxis .	45
6.1.2	Asymmetrische Information hinsichtlich der Qualität in der theoretischen Literatur zur Spitalregulierung . . . . .	48
6.2	Asymmetrische Information hinsichtlich der Behandlungsqualität im Modell mit doppeltem Unterstellungsverhältnis . . . . .	49
6.2.1	Annahmen bezüglich der Informationsverteilung . . . . .	49
6.2.2	First best-Allokation . . . . .	51
6.3	Analyse des Modells bei asymmetrischer Information und spitalinterner Prinzipal-Agent-Lösung . . . . .	52
6.3.1	Zeitliche Abfolge . . . . .	52
6.3.2	Bestimmung der Mengen $q_1$ und $q_2$ durch den Arzt . . . . .	53
6.3.2.1	Anpassung der Mengen an eine Budgetänderung . . . . .	54
6.3.3	Zahlungsbereitschaft des Arztes für eine Ausweitung des Budgets .	55

6.3.4	Bestimmung der Budgetierung $c^d(\theta)$ und der freien Mittel $t^d(\theta)$ durch das Management. . . . .	57
6.3.5	Anreizwirkung der Kostenbeteiligung $r(\theta)$ . . . . .	58
6.3.6	Optimale Pauschale $k(\theta)$ und Kostenbeteiligung $r(\theta)$ aus Sicht des Auftraggebers . . . . .	59
6.3.7	Bedeutungslosigkeit der spitalinternen Informationsasymmetrie . . .	64
6.4	Lösung des Modells bei spitalinterner Verhandlungslösung . . . . .	65
6.4.1	Zeitliche Abfolge . . . . .	65
6.4.2	Bestimmung der Mengen $q_1$ und $q_2$ durch den Arzt . . . . .	66
6.4.3	Bestimmung der Budgetierung $c^d(\theta)$ sowie der freien Mittel $t^d(\theta)$ durch Verhandlungen zwischen Management und Arzt . . . . .	66
6.5	Risikoaversion des Arztes gegenüber der freien Mittel . . . . .	72
6.5.1	First best-Lösung . . . . .	73
6.6	Lösung des Modells bei spitalinternem PA-Verhältnis und Risikoaversion des Arztes . . . . .	74
6.6.1	Bestimmung der Mengen $q_1$ und $q_2$ durch den Arzt . . . . .	74
6.6.2	Bestimmung der Budgetierung $c^d(\theta)$ sowie der freien Mittel $t^d(\theta)$ durch das Management . . . . .	74
6.6.3	Optimale Pauschale $k(\theta)$ und Kostenbeteiligung $r(\theta)$ aus Sicht des Auftraggebers . . . . .	76
6.7	Lösung des Modells bei spitalinterner Verhandlungslösung und Risikoaversion des Arztes . . . . .	76
6.7.1	Bestimmung der Budgetierung $c^d(\theta)$ sowie der freien Mittel $t^d(\theta)$ durch Verhandlungen zwischen Arzt und Management . . . . .	77
6.7.1.1	Komparative Statik . . . . .	78
6.7.2	Bedeutung der Verhandlungslösung für den Auftraggeber . . . . .	79
6.8	Fazit Kapitel 6 . . . . .	80
<b>7</b>	<b>Massnahmen des Managements zur Verbesserung der Kosteneffizienz</b>	<b>81</b>
7.1	Hintergrund . . . . .	81
7.2	Asymmetrische Information über die Anstrengungen zur Verbesserung der Kosteneffizienz in der Theorie der Spitalvergütung . . . . .	83
7.3	Massnahmen des Managements zur Verbesserung der Kosteneffizienz im Prinzipal-Agent-Modell mit doppeltem Unterstellungsverhältnis . . . . .	85
7.4	Asymmetrische Information hinsichtlich der Anstrengungen des Managements und des Preises $p_1$ im Modell mit doppeltem Unterstellungsverhältnis	87
7.4.1	Annahmen an die Informationsverteilung . . . . .	87
7.4.2	First best-Allokation . . . . .	87
7.5	Analyse des Modells bei spitalinterner Prinzipal-Agent-Lösung . . . . .	88

7.5.1	Zeitliche Abfolge . . . . .	88
7.5.2	Bestimmung der Mengen $q_1$ und $q_2$ durch den Arzt . . . . .	89
7.5.2.1	Anpassung der Mengen an eine Änderungen des Preises $p_1$ . . . . .	89
7.5.3	Kompensation des Arztes für eine Änderung des Preises $p_1$ . . . . .	90
7.5.4	Bestimmung des Preises $p_1$ , der Budgetierung $c^d(\theta)$ sowie der freien Mittel $t^d(\theta)$ durch das Management. . . . .	90
7.5.5	Maximierungsproblem des Auftraggebers: Trade-off zwischen Kosteneffizienz und Fallbudgetierung . . . . .	91
7.6	Analyse des Modells bei spitalinterner Verhandlungslösung . . . . .	93
7.6.1	Zeitliche Abfolge . . . . .	93
7.6.2	Bestimmung der Mengen $q_1$ und $q_2$ durch den Arzt . . . . .	94
7.6.3	Bestimmung der Budgetierung $c^d(\theta)$ sowie der freien Mittel $t^d(\theta)$ durch Verhandlungen zwischen Arzt und Management . . . . .	94
7.6.4	Bestimmung des Preises $p_1$ durch das Management . . . . .	96
7.6.5	Maximierungsproblem des Auftraggebers . . . . .	98
7.6.6	Die Kosten des Arbeitsleides und negativer Gewinn des Managements . . . . .	99
7.7	Fazit Kapitel 7 . . . . .	100
<b>8</b>	<b>Asymmetrische Information hinsichtlich der Fallschwere</b>	<b>102</b>
8.1	Asymmetrische Information hinsichtlich der Fallschwere zwischen Management und Auftraggeber . . . . .	103
8.1.1	Hintergrund . . . . .	103
8.1.1.1	Diagnosis Related Groups (DRG) - Abhängigkeit der Vergütung von der Fallschwere . . . . .	103
8.1.1.2	Bestimmung der Fallschwere im Spital . . . . .	104
8.1.1.3	Upcoding und Kontrolle . . . . .	105
8.1.2	Anreiztheoretischer Ansatz: Prinzip der direkten Offenbarung und Anreize zur wahrheitsgemässen Berichterstattung . . . . .	107
8.1.3	Asymmetrische Information hinsichtlich der Fallschwere zwischen Management und Auftraggeber im Prinzipal-Agent-Modell mit doppelten Unterstellungsverhältnis . . . . .	110
8.1.3.1	Annahmen bezüglich der Informationsverteilung . . . . .	110
8.1.3.2	Direktes Vergütungssystem . . . . .	110
8.1.3.3	Verteilung und Dichte der Fallschwere . . . . .	111
8.1.3.4	Optimale Allokation bei vollständiger Information über die Fallschwere . . . . .	111
8.1.4	Lösung des Modells bei spitalinterner Prinzipal-Agent-Lösung . . . . .	112
8.1.4.1	Zeitliche Abfolge . . . . .	112
8.1.4.2	Bestimmung der Mengen $q_1$ und $q_2$ durch den Arzt . . . . .	112

8.1.4.3	Anreizbedingung für wahrheitsgemässe Berichterstattung durch das Management . . . . .	113
8.1.4.4	Maximierungsproblem des Auftraggebers und optimale Allokation . . . . .	115
8.2	Spitalinterne Informationsasymmetrie hinsichtlich der Fallschwere . . . . .	118
8.2.1	Spitalinterne Informationsasymmetrie hinsichtlich der Fallschwere zwischen Arzt und Management im Prinzipal-Agent-Modell mit doppelem Unterstellungsverhältnis . . . . .	120
8.2.1.1	Informationsverteilung . . . . .	120
8.2.2	Lösung des Modells bei spitalinterner Prinzipal-Agent-Lösung . . . . .	120
8.2.2.1	Entscheidungen des Arztes: Bestimmung der Mengen $q_1$ und $q_2$ und Strategie der Berichterstattung . . . . .	120
8.2.2.2	Teilnahmebedingung des Arztes . . . . .	122
8.2.2.3	Entscheidungen des Managements: Berichterstattung an den Auftraggeber und Bestimmung der freien Mittel $t^d(\theta)$ . . . . .	122
8.2.2.4	Bedingung für wahrheitsgemässe Berichterstattung und Teilnahmebedingung des Managements . . . . .	124
8.2.2.5	Maximierungsproblem des Auftraggebers und optimale Allokation . . . . .	126
8.3	Relevanz der Teilnahmebedingungen . . . . .	128
8.3.1	Lösung des Modells bei asymmetrischer Information auf beiden Ebenen und bei bindender ex-ante Teilnahmebedingung . . . . .	132
8.4	Einfluss der Nash-Verhandlungslösung auf das Ergebnis . . . . .	137
8.4.1	Bedeutung der Verhandlungslösung bei vollständiger spitalinterner Information hinsichtlich der Fallschwere zwischen Arzt und Management . . . . .	139
8.4.2	Einfluss der Verhandlungslösung bei asymmetrischer Information hinsichtlich der Fallschwere zwischen Arzt und Management . . . . .	140
8.5	Fazit Kapitel 8 . . . . .	141
<b>9</b>	<b>Schlussfolgerung</b> . . . . .	<b>144</b>
9.1	Relevanz des spitalinternen Entscheidungsprozesses für das optimale Vergütungssystem . . . . .	144
9.2	Umsetzung des optimalen Systems: Nutzengewinn und Kosten für den Auftraggeber . . . . .	147
9.3	Beitrag zur Theorie der Spitalvergütung und Ausblick . . . . .	148
<b>A</b>	<b>Anhang</b> . . . . .	<b>152</b>
A.1	Unabhängigkeit der Mengen $q_1$ und $q_2$ von der Fallschwere . . . . .	152



A.2	Einfluss der Verhandlungsmacht und der Vergütungsparameter bei Risiko- aversion des Arztes . . . . .	153
A.2.1	Auswirkung der Verhandlungsmacht $w$ auf die freien Mittel $t^d(\theta)$ und die Budgetierung $c^d(\theta)$ . . . . .	153
A.2.2	Auswirkung der Kostenbeteiligung $r(\theta)$ und der Pauschale $k(\theta)$ auf die freien Mittel $t^d(\theta)$ und die Budgetierung $c^d(\theta)$ . . . . .	155
A.3	Das Arbeitsleid $\varphi$ als direkte Funktion des Preises $p_1$ . . . . .	158
A.4	Globale Bedingung für wahrheitsgemässe Berichterstattung des Managements	159
A.5	Globale Bedingung für wahrheitsgemässe Berichterstattung des Manage- ments bei spitalinterner asymmetrischer Information . . . . .	160
A.6	Nash-Verhandlungslösung bei spitalinterner asymmetrischer Information hinsichtlich der Fallschwere . . . . .	161
A.7	Symbole und Ausdrücke . . . . .	165

# Abbildungsverzeichnis

1	Modellübersicht . . . . .	25
2	Äquivalenz von Nash-Verhandlungslösung und Prinzipal-Agent-Lösung . .	71
3	Verlauf Fallbudgetierung und Nutzenüberschuss . . . . .	138

## Tabellenverzeichnis

1	Informationsverteilung des Modells in Kapitel 6. . . . .	51
2	Informationsverteilung des Modells in Kapitel 7. . . . .	87
3	Informationsverteilung des Modells in Abschnitt 8.1.3. . . . .	110
4	Informationsverteilung des Modells in Abschnitt 8.2.1. . . . .	121
5	Symbole und Ausdrücke . . . . .	165

# 1 Einleitung

## 1.1 Unvollständige Information auf dem Markt für Spitalleistungen

Der Markt für Spitalleistungen ist in der Schweiz massiven regulativen Eingriffen unterworfen. Der gewichtigste Eingriff findet beim Mechanismus der Preisbildung statt. Diese wird, wie in den meisten marktwirtschaftlichen Industrieländern, im Grossen und Ganzen nicht dem freien Markt überlassen, sondern ist meist das Ergebnis einer festgesetzten Berechnungsmethode.<sup>1</sup> Begründet werden die Eingriffe mit einem Versagen des Marktes für Gesundheitsleistungen, welches eine Pareto-effiziente Allokation als Gleichgewicht des unregulierten Marktes verhindert. Als Ursachen für das Marktversagen wiederum werden eine ganze Reihe von spezifischen Charakteristika genannt, die medizinischen Leistungen anhaften.<sup>2</sup> Ein häufig zitiertes Problem ist beispielsweise die *unvollkommene* Information hinsichtlich der Qualität von medizinischen Leistungen und die damit verbundene unvollständige Marktransparenz, welche einen funktionierenden Wettbewerb zwischen den Anbietern medizinischer Leistungen verhindert. Häufig besteht zum Beispiel ein unsicherer Zusammenhang zwischen der Qualität und Effektivität von medizinischen Leistungen, so dass aus einer Veränderung des Gesundheitszustandes nicht unbedingt auf die Qualität der Leistungen geschlossen werden kann. Die unvollständige Information erschwert es einem Konsumenten, die Qualität der medizinischen Leistungen zu beurteilen und mit jener der Leistungen anderer Anbieter zu vergleichen, um die Nachfrageentscheidung von der Qualität der angebotenen Leistung abhängig zu machen. Im extremsten Fall ist die Nachfrage nach den medizinischen Leistungen eines Anbieters, wie zum Beispiel eines Spitals, unabhängig von der bereitgestellten Qualität, so dass die Konkurrenz zwischen Spitälern nicht den Anreiz für das Spital enthält, die effiziente Qualität zu wählen. Zudem entsteht ein Problem der Kosteneffizienz. Bei perfekter Transparenz und vollkommenen Wettbewerb könnten nur die Spitäler überleben, die bei einer bestimmten Qualität mit dem entsprechenden Marktpreis für die Leistungen ihre Kosten zu decken in der Lage sind. Der Druck des perfekten Wettbewerbs enthielte also Anreize für das Spital, kosteneffizient zu behandeln. Besteht dagegen unvollständige Information und können die Konsumenten die Qualität des Spitals nicht beurteilen, ermöglicht dies dem Spital bei der Qualität zu sparen, um die Behandlung zu einem konkurrenzfähigen Preis anbieten zu können, ohne tatsächlich kosteneffizient zu sein.

---

<sup>1</sup>Einen Überblick über die Methodik der Preisbestimmung in vielen europäischen Ländern bieten Schreyögg et al. (2006).

<sup>2</sup>Eine umfassende Erörterung dieser Merkmale ist in Breyer/Zweifel/Kifmann (2005), Kapitel 5 zu finden.

## 1.2 Unzulänglichkeiten der regulativen Eingriffe in die Preisbildung

Ob hier allerdings die regulativen Eingriffe in die Preisbildung für Spitalleistungen, wie sie zum Beispiel in der Schweiz vorgenommenen werden, zu einer Allokation führen, die einem Gleichgewicht des unregulierten Marktes vorzuziehen ist, bleibt fraglich. Denn durch die Eingriffe wird das Problem der Qualitätsbereitstellung und kosteneffizienten Behandlung auch nicht gelöst. In der Schweiz sind die realisierten Behandlungskosten des einzelnen Spitals wesentlicher Ausgangspunkt der regulativen Festlegung des Preises, den es für die Behandlung erhält.<sup>3</sup> Damit hat auch ein ineffizientes Spital mehr oder weniger die Garantie, dass der Preis die Kosten deckt. Da nahezu alle medizinisch notwendigen Leistungen über eine Versicherung gedeckt sind (Kranken- oder Unfallversicherungen) und die versicherten Konsumenten die Kosten der Behandlung nur indirekt über die Prämien tragen, ist die Nachfrage sehr unelastisch hinsichtlich des Preises, d.h. ein vergleichsweise hoher Preis wirkt sich nicht oder kaum auf die Nachfrage des Spitals aus.<sup>4</sup> Zusätzlich gelten viele Spitäler aus gesundheitspolitischer Überlegung als unverzichtbar, so dass deren Bankrott mit entsprechenden staatlichen Defizitfinanzierungen verhindert wird. Unter diesen Voraussetzungen bestehen für das Spital nur sehr geringe Anreize, sich um eine Verbesserung der Effizienz zu bemühen. Eventuell mindert ein solches System die Anreize, an der Qualität zu sparen, jedoch resultiert daraus noch lange keine effiziente Bereitstellung. Beispielsweise ist es denkbar, dass ein Spital, das seine Kosten vollständig ersetzt bekommt, bei der Festlegung der Qualität ausschliesslich seine eigenen Präferenzen beachtet und einige Leistungen mit zu hoher, andere Leistungen dagegen mit zu niedrigerer Qualität erstellt.

Vom Eigeninteresse der regulierenden Instanz einmal abgesehen, welches auch nicht gezwungener Massen mit dem Erreichen einer effizienten Allokation im Einklang liegen muss, besteht bei regulativen Eingriffen das Problem, dass die für die Regulierung zur Verfügung stehenden Instrumente im Allgemeinen nicht alle relevante Grössen erfassen. Hier ist die unter dem Schlagwort *asymmetrisch verteilte Information* diskutierte Problematik ein zentraler Aspekt. Ein Zustand der asymmetrisch verteilten Information entsteht in diesem Zusammenhang, wenn der zu regulierende Leistungserbringer über Information bezüglich relevanter Grössen verfügt, die dem Regulator hingegen nicht zugänglich sind und welche

---

<sup>3</sup>Im Rahmen der neuen Spitalfinanzierung ab 2012 in der Schweiz wird die Abhängigkeit der spital-individuellen Preise von den jeweiligen Behandlungskosten dahingehend abgeschwächt, als dass die Behandlungskosten verschiedener Spitäler untereinander verglichen werden. Spitäler, die mit ihren Kosten oberhalb eines bestimmten Quantils der Gesamtheit der verglichenen Spitäler liegen, müssen mit Abschlägen rechnen (Benchmark-Verfahren).

<sup>4</sup>Hinzu kommt, dass zumindest vor 2012 eine sehr uneinheitliche Tarifstruktur die Vergleichbarkeit der Preise so gut wie unmöglich machte. Etwas mehr Transparenz schaffen ab 2012 die schweizweit einheitlich eingeführten SwissDRG- Fallpauschalen.

er deshalb nicht festlegen kann. Das oben genannte Problem der kosteneffizienten Behandlung bestünde nicht, wenn die für die Kosteneffizienz notwendigen Anstrengungen des Spitals beobachtbar wären und damit regulatorisch festgelegt werden könnten.

Wenn bestimmte Grössen nicht festzulegen sind, bedeutet dies jedoch nicht, dass der Regulator überhaupt keinen Einfluss auf das Spital hinsichtlich dieser Grössen nehmen kann. Mit den Eingriffen sind immer auch *Anreize* verbunden, die - ob bewusst oder unbewusst eingesetzt - das Verhalten des Spitals beeinflussen. Beispielsweise enthält ein System, welches dem Spital einen fixen Betrag zur Behandlung zur Verfügung stellt, Anreize für das Spital, die Leistungen kosteneffizient zu erbringen. Gleichzeitig gehen von einem solchen „prospektiven“ Vergütungssystem allerdings auch die unerwünschten Anreize aus, nämlich die Qualität zu senken oder die Patienten nach ihrer finanziellen Vorteilhaftigkeit zu selektieren. Ein vorgenommener regulatorischer Eingriff, bei dem die Auswirkungen der von ihm ausgehenden Anreize ignoriert oder falsch eingeschätzt werden, wird nicht zum gewünschten Ergebnis führen und eine aus Sicht des Regulators suboptimale Allokation zur Folge haben. Eine optimale Regulierung beinhaltet daher immer die richtigen Anreize, die der regulatorische Eingriff dem Leistungserbringer setzt.

### 1.3 Das Prinzipal-Agent-Modell in der ökonomischen Theorie der Regulierung

In der ökonomischen Theorie der Regulierung sind asymmetrische Information und Anreize bei der Regulierung von natürlichen Monopolen und öffentlichen Betrieben seit Anfang der 1980er Jahre zentrale Punkte der Diskussion [Laffont/Tirole (1993)]. Als Rahmen der Analyse der optimalen Regulierung von Firmen hat sich dabei das *Prinzipal-Agent-Modell* (im Folgenden PA-Modell) etabliert. Die zentralen Annahmen dieses Modells in seiner einfachsten Form sind:

- Ein Prinzipal *delegiert* die Erfüllung eines Auftrages an einen Agenten.
- Die Modalitäten der Auftragsdurchführung, insbesondere die Vergütung des Agenten, werden in einem *Vertrag* geregelt, den der Prinzipal entwirft und dem Agenten als *take-it-or-leave-it* Angebot unterbreitet.
- Der Agent seinerseits verfügt hinsichtlich der Ausprägung von zur Auftragsdurchführung entscheidender Grössen über *private Informationen*, d.h. der Prinzipal kann diese Grössen nicht beobachten und nicht mit dem Vertrag regeln.
- Die Interessen des Agenten sind nicht *deckungsgleich* mit jenen des Auftraggebers.

Die aus diesen Annahmen resultierende Problematik besteht darin, dass der Prinzipal mit der Delegation der Auftragsdurchführung die Entscheidung über relevante Grössen dort dem Agenten überlassen muss, wo die Ausprägung der Grössen private Information des Agenten ist und eine Überprüfung bzw. vertragliche Regelung nicht möglich ist. Aus Sicht des Prinzipals führt dies zu Problemen, da die Interessen des Agenten nicht deckungsgleich mit den eigenen sind. Durch die fehlende Möglichkeit des Auftraggebers, alle relevanten Grössen vertraglich zu regeln, entstehen dem Agenten bei der Durchführung des Auftrages Freiräume, die eigenen Interessen zu verfolgen. Sind die Interessen von Agent und Prinzipal nicht deckungsgleich, geht es zulasten des Prinzipals, wenn der Agent diese Freiräume ausnutzt.

Das PA-Modell bietet den Rahmen zur theoretischen Analyse, wie der Prinzipal in optimaler Weise mit dem beschriebenen Problem umgeht. In erster Linie geht es dabei um Anreize, die dem Agenten über die Vergütung gesetzt werden und ihn dazu veranlassen, sich soweit wie möglich im Sinne des Prinzipals zu verhalten. Im Sinne des Prinzipals bedeutet in diesem Zusammenhang, dass aus der Interaktion zwischen Prinzipal und Agent eine aus Sicht des Prinzipals bestmögliche *Allokation* resultiert.

## 2 Fragestellung und Gliederung

### 2.1 Fragestellung

Auch in der theoretischen Analyse der Spitalregulierung findet der Rahmen des PA-Modells in zahlreichen Artikeln Anwendung. Im Kontext der Erbringung von Spitalleistungen delegiert dabei ein Auftraggeber als Prinzipal die Behandlung eines oder mehrerer Patienten an das Spital als Agenten. Mit der Delegation überlässt der Auftraggeber die Entscheidung über die zur Behandlung eingesetzten Ressourcen und Massnahmen wie Diagnostik, Wahl der Therapie und Pflege dem Spital. Das Spital verfügt über private Informationen hinsichtlich mehrerer für die Durchführung der Behandlung relevanter Grössen. Wichtigste Beispiele für diese Grössen sind die Qualität der Behandlung [z.B. Jack (2005), Eggleston (2005), Chalkley/Malcomson (1998a+1998b)], Anstrengungen zur Verbesserung der Kosteneffizienz [z.B. Mougeot (2009), Jack (2005), Chalkley/Malcomson (2002), De Fraja (2000), Chalkley/Malcomson (1998a+1998b)] sowie die Fallschwere der behandelten Patienten [z.B. Mougeot (2009), Chalkley/Malcomson (2002)].

Die asymmetrische Information allein führt allerdings noch zu keinen Problemen für den Auftraggeber. Erst in Kombination mit der Divergenz der Ziele von Spital und Auftraggeber resultiert für den Auftraggeber das Problem, dass sich das Spital bei den vertraglich

nicht zu fassenden Grössen nicht in seinem Sinne verhält. Hätten Auftraggeber und Spital bei allen Aspekten übereinstimmende Interessen, könnte der Auftraggeber das Spital für die Durchführung des Auftrages lediglich mit einer fixen Pauschale vergüten und trotzdem die first best-Allokation erreichen. Weitere vertragliche Regelungen oder Anreize wären nicht notwendig, da sich das Spital von sich aus ganz im Sinne des Auftraggebers verhalten würde.

In den meisten Artikeln, in denen Fragen der Spitalregulierung im Rahmen des PA-Modells analysiert werden, resultieren die divergierenden Interessen von Spital und Auftraggeber aus Unterschieden in den von den Autoren *postulierten* Zielen von Spital und Auftraggeber. Da das Spital eine Organisation ist, deren Verhalten von mehreren spitalinternen Entscheidungsträgern bestimmt wird, steht hinter dem Postulat der Ziele die implizite Annahme der Existenz eines - wie auch immer gearteten - Entscheidungsprozesses, der die idiosynkratischen Präferenzen der einzelnen Entscheidungsträger derart zusammenführt, dass für das Spital als ganzes eben die postulierten Ziele resultieren.

In der vorliegenden Arbeit soll hinsichtlich der Ziele des Spitals der umgekehrte Weg gewählt werden. Anstatt die Ziele des Spitals als angenommenes Ergebnis eines nicht näher bekannten Entscheidungsprozesses zu postulieren, sollen ausgehend von den Präferenzen der wichtigsten Entscheidungsträger mit der expliziten Modellierung eines spitalinternen Entscheidungsprozesses die Ziele als Ergebnis dieses Entscheidungsprozesses hergeleitet werden. Dieser Ansatz ermöglicht es nicht nur, den Zielen des Spitals im PA-Modell eine gewisse formale Begründung zu geben, sondern auch die Ergebnisse unterschiedlicher Entscheidungsprozesse miteinander zu vergleichen.

Für einen Auftraggeber hat diese Analyse bei der Entwicklung eines optimalen Anreizsystems durchaus Relevanz. So sind die Reaktionen des Spitals auf die Anreize eines Vergütungssystems eine direkte Implikation der Ziele, die das Spital bei der Auftragsdurchführung verfolgt. Wenn aus verschiedenen spitalinternen Entscheidungsprozessen unterschiedliche Zielsetzungen des Spitals folgen, ist es für den Auftraggeber von grosser Bedeutung, über die spitalinternen Gegebenheiten und deren Implikationen hinsichtlich der Zielsetzung des Spitals informiert zu sein. Geht der Auftraggeber zum Beispiel von einer Zielsetzung des Spitals aus, die bei dem im Spital implementierten Entscheidungsprozess gar nicht möglich ist, haben die Anreize des Vergütungssystems natürlich auch nicht die vom Auftraggeber beabsichtigte Wirkung.

Zur Analyse der Auswirkung des spitalinternen Entscheidungsprozesses auf die Zielsetzung des Spitals und auf das optimale Anreizsystem des Auftraggebers wird in dieser Arbeit ein einfaches PA-Modell durch eine explizite Modellierung des spitalinternen Ent-



scheidungsprozesses als ein nicht-kooperatives Spiel zwischen Chefarzt und Spitalleitung bzw. Management erweitert, wobei das Prinzipal-Agent-Verhältnis zwischen Auftraggeber und Spital (als Organisation mit den Entscheidungsträgern Management und Arzt) erhalten bleibt.

Konkret werden zwei Entscheidungsmechanismen betrachtet und verglichen: Der erste Mechanismus besteht in einem weiteren Prinzipal-Agent-Verhältnis, wobei der Arzt als Agent des Managements die Behandlung von Patienten durchführt, während dem Management eine Doppelrolle zukommt: Zum einen fungiert es als Prinzipal des Arztes, zum anderen als Agent des Auftraggebers.

Der zweite Mechanismus besteht in einer Verhandlungslösung zwischen Arzt und Management. Auch bei diesem Mechanismus besteht das Prinzipal-Agent-Verhältnis zwischen Auftraggeber und Management. Jedoch muss das Management spitalintern das Verhandlungsergebnis mit dem Arzt beachten. Um die hierarchische Struktur des Modell herauszustellen, wird das Modell im Folgenden, wie dem Titel dieser Arbeit bereits zu entnehmen ist, als „Prinzipal-Agent-Modell mit doppeltem Unterstellungsverhältnis“ bezeichnet.

Die Analyse findet hinsichtlich der asymmetrisch verteilten Information unter verschiedenen Szenarien statt, wobei sich die Informationsasymmetrie auf die Qualität der Behandlung, spitalinterne Massnahmen zur Verbesserung der Kosteneffizienz und die Schwere der Erkrankung des Patienten bezieht.

## 2.2 Gliederung

Die weitere Gliederung der Arbeit ist wie folgt: In Kapitel 3 werden zunächst einige wichtige Aspekte des PA-Modells dargelegt. Kapitel 4 beginnt mit einer Betrachtung der divergierenden Ziele von Auftraggeber und Spital in einem häufig zitierten Modell von Chalkley und Malcomson, worauf eine Diskussion der Begründung der Zielfunktionen folgt. Um die Motivation zu untermauern, den spitalinternen Entscheidungsprozess explizit zu modellieren, wird am Ende des Kapitels anhand eines kurzen Beispiels demonstriert, welchen Einfluss eine geringfügige Variation eines spitalinternen Entscheidungsprozesses auf die Ziele des Spitals haben kann. In Kapitel 5 erfolgt eine Erläuterung der grundsätzlichen Annahmen des hier entwickelten erweiterten PA-Modells mit doppeltem Unterstellungsverhältnis. Die eigentliche Analyse des Modells findet in den Kapiteln 6 – 8 statt, wobei sich die Analysen in den einzelnen Kapiteln vor allem hinsichtlich der Annahmen an die asymmetrisch verteilte Information unterscheiden. In Kapitel 6 wird angenommen, die zur Behandlung des Patienten eingesetzten Mengen der medizinischen Leistungen sind private Informationen des Arztes und vom Management und vom Auftraggeber vertraglich

nicht festzulegen. Die Lösungen des Modells unter diesen Voraussetzungen bei internem Prinzipal-Agent-Verhältnis werden mit der Lösung bei interner Verhandlungslösung verglichen. Die Analyse in Kapitel 7 erfolgt unter der Annahme, dass zusätzlich zu den Mengen der medizinischen Leistungen die Anstrengungen des Managements zur Verbesserung der Kosteneffizienz für den Auftraggeber nicht zu beobachten und nicht vertraglich festzulegen sind. Auch hier wird die Lösung bei spitalinternem Prinzipal-Agent-Verhältnis und die Verhandlungslösung ermittelt. In Kapitel 8 dreht sich die Analyse um die asymmetrische Verteilung von Informationen bezüglich der Fallschwere. Hier liegt der Fokus vor allem auf dem Einfluss der spitalinternen asymmetrischen Information zwischen Arzt und Management. In Kapitel 9 schliesslich werden die Ergebnisse zusammengefasst und ein Fazit aus der Analyse gezogen.

## 3 Einführung in das Prinzipal-Agent-Modell

### 3.1 Einordnung des Prinzipal-Agent-Modells in die Anreiztheorie

Das PA-Modell kommt zu Analysen in verschiedensten ökonomischen Bereichen zur Anwendung, allerdings bleibt die Grundproblematik diejenige, die bereits in der Einleitung dargestellt wurde: Ein Prinzipal überträgt die Durchführung eines Auftrages an einen Agenten. Hierfür entwirft er einen Vertrag, den er dem Agenten als take-it-or-leave-it Angebot unterbreitet. Für den Prinzipal ist die Delegation der Durchführung problematisch, da der Agent über private Information bezüglich relevanter Grössen verfügt *und* bei der Durchführung andere Ziele verfolgt als der Prinzipal. Da der Prinzipal wegen der asymmetrischen Verteilung von Informationen nicht alle relevanten Grössen vertraglich regeln kann, bestehen für den Agenten Freiräume bei der Ausführung des Auftrages. Wenn der Agent dabei Ziele verfolgt, die nicht mit den Zielen des Prinzipals übereinstimmen, resultiert eine Allokation, die von jener abweicht, die der Prinzipal wählen würde, wenn er alle relevanten Grössen vertraglich festlegen könnte.

Die optimale Gestaltung des Vertrages und der darin enthaltenen Anreize im PA-Modell ist thematisch der *Anreiztheorie* bzw. dem *Mechanismus-Design* zuzuordnen. In dieser Theorie führen ein oder mehrere Individuen (Agenten) für andere nicht beobachtbare Aktionen aus oder sind Träger von Informationen, zu denen andere keinen Zugang haben. Von den Aktionen und Informationen der einzelnen Agenten hängt die Zielfunktion eines Prinzipals ab. Im Zentrum der Anreiztheorie steht nun die Maximierung der Zielfunktion des Prinzipals mit dem Design eines optimalen Mechanismus bzw. Vertrages, der das Verhalten der Agenten in dieser Hinsicht optimal koordiniert. Das Problem dabei ist, dass der Mechanismus nicht oder nicht ohne weiteres auf die Aktionen oder Informationen abstellen kann, da diese nur vom jeweiligen Agenten beobachtet werden können. Stattdessen muss über die Ausgestaltung von Anreizen Einfluss auf die Agenten genommen werden.

Innerhalb der Anreiztheorie bildet das einfache PA-Modell eine Situation ab, in der ein Prinzipal die Durchführung eines Auftrages an einen Agenten delegiert. Der Agent hat im PA-Modell gewissermassen eine Monopolstellung inne, da kein Wettbewerb mit anderen Agenten stattfindet. Das Element des Wettbewerbes spielt dagegen bei der Anwendung der Anreiztheorie auf dem Gebiet der Auktionstheorie eine zentrale Rolle. Hier macht sich der Prinzipal zunutze, dass mehrere Agenten um eine Sache bieten und damit eher dazu bereit sind, ihren Informationsstand preiszugeben. Ein weiteres Beispiel für eine Anwendung der Anreiztheorie dreht sich um die Bereitstellung eines öffentlichen Gutes, wenn die Präferenzen der einzelnen Konsumenten der über die Allokation des Gutes entscheidenden Behörde nicht bekannt sind. Hier stellt sich für die Behörde als Prinzipal die Frage, ob

das Gut beschafft werden soll, und wie die einzelnen Konsumenten als Agenten an der Finanzierung beteiligt werden sollen.

## 3.2 Struktur des Prinzipal-Agent-Modells

Wichtige Merkmale der Struktur des PA-Modells sind die *Zielfunktionen* des Prinzipals und des Agenten, die *Anreizverträglichkeitsbedingungen*, die *Teilnahmebedingung* des Agenten sowie der *zeitliche Ablauf* des Modells. Da im Laufe der Analyse in dieser Arbeit häufig auf diese Aspekte Bezug genommen wird, werden sie im Folgenden kurz erläutert.

### 3.2.1 Zielfunktionen

Mit den Zielfunktionen werden die Präferenzen des Prinzipals und des Agenten hinsichtlich der aus der Interaktion resultierenden Allokation beschrieben. Sei zum Beispiel  $\{t, A\}$  diese Allokation, wobei  $t \in T \subset \mathbb{R}$  die Vergütung des Agenten durch den Prinzipal für die Durchführung des Auftrages ist, und  $A = (a_1, \dots, a_h)$  mit  $a_i \in A_i$  ein Vektor mit weiteren relevanten Aspekten der Allokation, wie zum Beispiel dem Output aus der Auftragsdurchführung. Weiter sei  $\Theta = (\theta_1, \dots, \theta_l)$  mit  $\theta_i \in \Theta_i \subset \mathbb{R}$  eine Realisation eines Vektors mit exogen gegebenen Grössen. Die Nutzenfunktion des Prinzipals ordnet jeder möglichen Allokation  $\{t, A\}$  für jede Realisation  $\Theta$  eine reelle Zahl  $W$  zu:

$$W = W(t, A, \Theta) \in \mathbb{R}. \quad (1)$$

Für die Zielfunktion des Agenten gilt entsprechend

$$U = U(t, A, \Theta) \in \mathbb{R}. \quad (2)$$

In der Regel sind  $W(t, c, A, \Theta)$  und  $U(t, A, \Theta)$  *Bernoulli-Nutzenfunktionen*, d.h. bei Unsicherheit lässt sich der erwartete Nutzen von Prinzipal bzw. Agent mit dem Erwartungswert von (1) bzw. (2) darstellen [Mas-Colell/Whinston/Green (1995), Kapitel 8].<sup>5</sup>

Aus den im Modell postulierten Zielfunktionen resultieren die Divergenzen der Ziele zwischen Prinzipal und Agent.

---

<sup>5</sup>Häufig wird ein stochastischer Zusammenhang zwischen dem Arbeitseinsatz des Agenten und dem Ergebnis der Auftragsdurchführung angenommen [z.B. Holmstrom/Milgrom (1991)]. Unsicherheit ergibt sich auch hinsichtlich der exogenen Grössen, wenn Entscheidungen getroffen werden, bevor die tatsächliche Ausprägung der Grössen bekannt wird.

### 3.2.2 Anreizverträglichkeitsbedingungen und Arten der asymmetrischen Information

Eine Allokation, die als Lösung des PA-Modells folgt, muss in den vertraglich nicht regelbaren Bereichen die Ausprägungen enthalten, die der Agent freiwillig wählt, da der Agent seine Freiräume zur Maximierung seiner eigenen Zielfunktion ausnutzt. Eine bestimmte Allokation ist für den Prinzipal also nur dann zu erreichen, wenn er in den nicht regelbaren Bereichen dem Agenten Anreize setzen kann, die relevanten Grössen der gewünschten Allokation entsprechend zu wählen. Dafür ist es notwendig, dass der Prinzipal bei der Ausgestaltung des Vertrages die *Anreizverträglichkeitsbedingungen* des Agenten beachtet. Die Form der Anreizverträglichkeitsbedingungen hängt grundsätzlich davon ab, ob die asymmetrisch verteilte Information bei Grössen vorliegt, die vom Agenten gewählt werden (versteckte Handlung), oder Grössen betrifft, die exogen gegeben sind (versteckte Information).

**3.2.2.1 Anreizverträglichkeitsbedingung bei versteckter Handlung** Bezieht sich die zwischen Agent und Prinzipal asymmetrisch verteilte Information auf relevante Grössen, deren Ausprägung der Agent bestimmen kann, nachdem er dem Vertragsangebot des Prinzipal zugestimmt hat, führt dies potentiell zu einer Problematik der *verborgenen Handlung* [Breyer/Kolmar (2005), Seite 307] (oft als *Moral Hazard* bezeichnet). Sei  $\mathbf{q} = (q_1, \dots, q_n)$  mit  $q_i \in Q_i \subset \mathbb{R}$ ,  $i = 1, \dots, n$  ein Vektor mit relevanten Grössen, die der Agent unbeobachtet vom Prinzipal wählen kann. Die Bestimmung von  $\mathbf{q}$  führe zum Output  $A = A(\mathbf{q})$  und ziehe die Kosten der Auftragsdurchführung  $c = c(\mathbf{q})$  nach sich. Ferner sei angenommen, der Prinzipal könne die Kosten beobachten und als Argument für die Entlohnung  $t = t(c)$  des Agenten als Argument verwenden. Bei gegebenem  $\Theta$  ist eine bestimmte Allokation  $\{t, A(\mathbf{q}^*)\}$  nur dann mit dem Vertrag  $\{t(c)\}$  möglich, wenn  $\mathbf{q}^*$  den Nutzen des Agenten maximiert. Die Anreizbedingung bei versteckter Handlung lautet dann

$$\mathbf{q}^* \in \arg \max_{\mathbf{q}} U(t(c(\mathbf{q})), A(\mathbf{q}), \Theta). \quad (3)$$

**3.2.2.2 Anreizverträglichkeitsbedingung bei versteckter Information** Ein anderer Fall der asymmetrischen Information ergibt sich, wenn exogen gegebene Grössen nur vom Agenten zu beobachten sind. Zum Problem wird diese *versteckten Information*, wenn der Agent in Kenntnis dieser Grössen Entscheidungen fällen und aus dem Informationsvorsprung gegenüber dem Prinzipal eine *Informationsrente* ziehen kann.<sup>6</sup>

---

<sup>6</sup>Dieser Fall wird auch als Antiselektion oder ‚Adverse Selection‘ bezeichnet.

In der Regel wird der Prinzipal die im Vertrag geregelten Konditionen der Auftragsdurchführung und die Vergütung des Agenten vom jeweiligen Zustand exogen gegebener Grössen abhängig machen. Da der Prinzipal die Ausprägung der Grössen aber nicht kennt, bleibt ihm dabei allerdings nichts anderes übrig, als den Agenten direkt oder indirekt nach der Ausprägung zu fragen. Der Agent wird dabei jedoch nicht unbedingt die tatsächliche Ausprägung bekannt geben, sondern den Wert, der bei den vom Prinzipal definierten Vertragsmodalitäten seinen Nutzen maximiert. Bei asymmetrischer Information bezüglich einer relevanten exogen gegebenen Grösse ist eine bestimmte Allokation für den Prinzipal nur erreichbar, wenn sie mit diesem Verhalten des Agenten vereinbar ist, d.h. wenn die entsprechende Anreizverträglichkeitsbedingung erfüllt ist.

Sei zur Vereinfachung  $\Theta$  ein eindimensionaler Vektor und  $\theta \in \mathbb{R}$  die tatsächliche Realisation der exogenen Grösse, die nur vom Agenten zu beobachten ist. Zur weiteren Vereinfachung sei angenommen, der Prinzipal könne die Aktionen des Agenten  $\mathbf{q}$  beobachten und vertraglich festlegen. Weiter nehmen wir an, der Prinzipal biete dem Agenten einen Vertrag an, der die Vergütung und die festgelegte Vorgabe  $\mathbf{q}$  direkt von der vom Agenten offenbarten Ausprägung der exogenen Grösse abhängig macht.<sup>7</sup> Sei  $\{t(\theta), \mathbf{q}^*(\theta)\} \forall \theta \in \Theta$  dieser Vertrag. Eine Allokation  $\{t^*(\theta), A(\mathbf{q}^*(\theta))\}$  ist bei einer Realisation  $\theta$  mit dem Vertragsmenü nur dann implementierbar, wenn die Anreizbedingung

$$\theta \in \arg \max_{\hat{\theta}} U(\theta, t^*(\hat{\theta}), A(\mathbf{q}^*(\hat{\theta})), \theta) \quad (4)$$

erfüllt ist, wobei  $\hat{\theta} \in \Theta$  die vom Agenten offenbarte Ausprägung der exogenen Grösse ist.

### 3.2.3 Teilnahmebedingung

Der Agent kann das Vertragsangebot des Prinzipals ablehnen. Da es sich aber um ein take-it-or-leave-it Angebot handelt, ist es ihm im Rahmen des PA-Modells nicht möglich, dem Prinzipal ein Gegenangebot zu unterbreiten. Wenn der Agent ablehnt, bleibt ihm also nur eine zum Auftrag des Prinzipals alternative Beschäftigung. Bei seiner Entscheidung über Annahme oder Ablehnung des Angebotes des Auftraggebers vergleicht der Agent den Nutzen, den er mit der Durchführung des Auftrages und der damit verbundenen Vergütung erhält, mit dem Nutzen, den ihm die beste alternative Betätigung bietet. Liegt der Wert des Angebotes unter dem als *Reservationsnutzen* bezeichneten Wert der besten Alternative, lehnt der Agent das Vertragsangebot ab. Vom Reservationsnutzen wird normalerweise angenommen, dass er exogen gegeben ist. Bei seinem Vertragsangebot muss

---

<sup>7</sup>Ein solcher Vertrag wird auch als *direkter* Mechanismus bezeichnet (vgl. Abschnitt 8.1.2 dieser Arbeit).

der Prinzipal neben den Anreizverträglichkeitsbedingungen also auch die *Teilnahmebedingung* des Agenten beachten und dafür Sorge tragen, dass aus Sicht des Agenten der Wert des Angebotes mindestens seinem Reservationsnutzen entspricht.

### 3.2.4 Zeitlicher Ablauf

Das PA-Modell ist ein mehrstufiges Spiel, dessen Ablauf der Spielstufen sich auf das Ergebnis auswirkt. Typischerweise sind folgende Ereignisse im PA-Modell entscheidend: Das Angebot und die Entscheidung des Agenten über Annahme oder Ablehnung desselben, der Zeitpunkt, an dem der Agenten die Ausprägung der exogenen Grössen erfährt, die Ausführung des Auftrages und die Festlegung der dafür notwendigen Grössen durch den Agenten, die Realisation des Ergebnisses und die Vergütung. Die zeitliche Reihenfolge dieser Ereignisse variiert von Anwendung zu Anwendung des Modells. Das Grundmodell in Lehrbuch von Laffont und Tirole hat zum Beispiel folgenden zeitlichen Ablauf [Laffont/Tirole (1993), Kapitel 1].

1. Nur der Agent erfährt die Ausprägung der exogenen Variablen.
2. Der Prinzipal bietet dem Agenten den Vertrag an.
3. Der Agent entscheidet über die Teilnahme am Vertrag.
4. Wenn der Agent dem Angebot zustimmt, erbringt er die im Zusammenhang der mit Auftragsdurchführung notwendigen Leistungen.
5. Die Kosten der Auftragsdurchführung werden realisiert, die Vergütung des Agenten findet statt.

Die Reihenfolge der Ereignisse hat Einfluss auf das Ergebnis des Modells. Eine besondere Bedeutung kommt hierbei dem Zeitpunkt zu, an dem der Agent die Ausprägung der exogenen Variablen erfährt. Geschieht dies, bevor er eigene Entscheidungen treffen muss, kann er seine Entscheidungen unter Sicherheit fällen. Muss er jedoch die Entscheidungen treffen, bevor er die Ausprägung in Erfahrung bringt, fällt er seine Entscheidungen entsprechend den Erwartungen über die jeweilige Realisation [vgl. Laffont/Martimort (2002), Kapitel 7].<sup>8</sup>

---

<sup>8</sup>In Abschnitt 8.3 wird näher auf die Frage eingegangen, welchen Einfluss es hat, wenn der Agent die Ausprägung einer exogenen Grösse vor oder nach der Entscheidung über seine Teilnahme am Vertrag erfährt.

### 3.3 Ergebnis des Prinzipal-Agent-Modells

Die Lösung des Modells charakterisiert die Allokation, welche die Zielfunktion des Prinzipals unter Einhaltung der Anreizverträglichkeitsbedingungen und der Teilnahmebedingung des Agenten maximiert.<sup>9</sup> Dabei definieren die Anreizbedingungen und die Teilnahmebedingung des Agenten die Menge der *erreichbaren* Allokationen [vgl. Laffont/Martimort (2002), Seite 30], die nur eine Teilmenge der Menge aller *möglichen* Allokationen  $T \times A$  ist. Mit der entsprechenden Ausgestaltung seines Vertrages gibt der Prinzipal dem Agenten Anreize, bei gegebenem  $\theta$  den Vektor  $\mathbf{q}$  derart zu bestimmen, dass die für den Prinzipal optimale Allokation unter den erreichbaren Allokationen resultiert. Wenn es ihm dabei nicht gelingt, mit der Ausgestaltung der Anreize die Interessen des Agenten den eigenen anzugleichen und daher Divergenzen der Ziele bestehen bleiben, dann schränken ihn die Anreizbedingungen und die Teilnahmebedingung des Agenten ein. In diesem Fall ist die Allokation, die er unter *vollständiger* Information erreichen könnte (meist als *first best*-Allokation bezeichnet), kein Element der Menge der erreichbaren Allokationen, so dass die best-möglich erreichbare Allokation nur eine *second best*-Lösung darstellt.

---

<sup>9</sup>Die Anreizverträglichkeitsbedingungen entsprechen den Reaktionsfunktionen eines Stackelberg-Folgers in einem Stackelberg-Spiel. Da der Prinzipal bei der Bestimmung des optimalen Vertrages mit den Anreizverträglichkeitsbedingungen die Reaktionen des Agenten auf den Vertrag beachten muss, wird der Prinzipal auch bisweilen als Stackelberg-Anführer bezeichnet.



## 4 Divergenzen der Ziele von Spital und Auftraggeber sowie Begründung der Zielfunktion des Spitals

Nachdem im vorigen Kapitel die Grundzüge des PA-Modells dargelegt wurden, ist es Ziel dieses Kapitels, mit der Diskussion eines häufig zitierten Modells von Chalkley und Malcomson aus dem Jahre 1998 [Chalkley/Malcomson (1998a)] einen Eindruck der Anwendung des PA-Modells auf dem Gebiet der Spitalregulierung zu vermitteln, wobei der Fokus auf den postulierten Zielfunktionen und den daraus folgenden Divergenzen der Ziele liegt.

### 4.1 Zielfunktionen und Divergenzen der Ziele von Auftraggeber und Spital im Modell von Chalkley und Malcomson

In dem Modell von Chalkley und Malcomson delegiert ein Auftraggeber als Prinzipal die Behandlung mehrerer Patienten einer bestimmten Fallschwere an das Spital als Agenten. Exogene Grössen spielen in dem Modell keine Rolle, die asymmetrische Information resultiert also in einer Problematik der versteckten Handlung. Mit ihrem Modell untersuchen Chalkley und Malcomson die Bedeutung der altruistischen Einstellung des Spitals für den Auftraggeber bei der optimalen Vergütung des Spitals, d.h. in wie weit sich der Auftraggeber eventuelle altruistische Motivationen des Spitals zunutze machen kann, um eine bestmögliche Allokation zu erreichen.

#### 4.1.1 Zielfunktion des Auftraggebers

Wie die meisten anderen Autoren nehmen Chalkley und Malcomson an, der Auftraggeber verkörpere eine Institution, die gesamtgesellschaftliche Ziele verfolgt und eine utilitaristische Wohlfahrtsfunktion maximiert, in der auch der Nutzen des Spitals  $S(\cdot)$  ungewichtet enthalten ist. Die Funktion hat die Form

$$W(e, q, n, t) = B(q, n) + S(e, q, n, t) - (1 + \alpha)(t + c(e, q, n)) \quad (5)$$

wobei  $q$  die Behandlungsqualität,  $n$  die Anzahl der behandelten Patienten und  $t$  die Vergütung des Spitals darstellt. Der Parameter  $e$  bezeichnet die Anstrengung des Spitals, die Kosteneffizienz zu verbessern. Der Ausdruck  $B(q, n)$  ist der Nutzen des Auftraggebers bzw. der Gesellschaft aus der Behandlung.  $c(e, q, n)$  sind die Kosten der Behandlung, die mit den Anstrengungen  $e$  des Spitals, die Effizienz zu verbessern, sinken, aber mit der Behandlungsqualität und der Anzahl der Patienten steigen.<sup>10</sup>  $\alpha \geq 0$  stellt die Schattenkosten der Finanzierung dar. Häufig wird angenommen, für die finanziellen Aufwendungen,

---

<sup>10</sup>In dieser Formulierung trägt der Auftraggeber die Verantwortung für die Kosten der Behandlung.

bestehend aus den Behandlungskosten und der Vergütung des Spitals, werden über obligatorische Abgaben eingenommene Gelder verwendet (Steuern, obligatorische Krankenversicherungsbeiträge), die aufgrund der verzerrenden Wirkung von Steuern und anderen „Reibungsverlusten“ für die Gesellschaft mit zusätzlichen Schattenkosten behaftet sind.<sup>11</sup>

#### 4.1.2 Zielfunktion des Spitals

Chalkley und Malcomson postulieren

$$S(e, q, n, t) = V(q, n) + t - \varphi(e, q, n), \quad (6)$$

als Zielfunktion des Spitals, wobei  $V(q, n)$  den aus der altruistischen Motivation entspringende Nutzen des Spitals aus der Behandlung darstellt, der in beiden Argumenten steigt. Ein in diesem Sinne altruistisches Spital ist prinzipiell bereit, Qualität und Quantität über das Niveau hinaus bereitzustellen, bei dem der marginale finanzielle Überschuss gleich null ist, solange dem Patienten dadurch zusätzlicher Nutzen entsteht.<sup>12</sup>  $\varphi(e, q, n)$  stellt den Nutzenverlust des Spitals aus dem Arbeitsleid dar, das den im Spital tätigen Mitarbeiter durch die physischen und psychischen Belastungen bei Durchführung des Auftrages entsteht.

In der Literatur wird eine solche oder ähnliche altruistische Komponente zum Beispiel auch von Mougeot/Naegele (2009), Jack (2005), Eggleston (2005) oder Ellis (1998) postuliert. Dagegen gehen Fees/Ossig (2007), Siciliani (2006) oder Ma (1994) von einem rein am finanziellen Gewinn orientierten Spital aus.

#### 4.1.3 Optimale Anreize und Divergenzen der Ziele

**4.1.3.1 Bedingungen für die first best-Allokation** Mit der Nutzenfunktion des Spitals (6) ergibt sich für die Zielfunktion (5) des Auftraggebers bei Chalkley und Malcomson<sup>13</sup>

---

<sup>11</sup>Ballard, Shoven und Whalley schätzen, dass der zusätzliche Schattenpreis steuerfinanzierter Projekte in den USA zwischen 15% und 50% liegt [Ballard/Shoven/Whalley (1985)].

<sup>12</sup>Die altruistische Komponente lässt sich in der Realität vermuten, wenn zum Beispiel Patienten behandelt werden, bei denen die Wahrscheinlichkeit eines Zahlungsausfalles sehr hoch ist. In der Praxis ist eine solche 'charity care' tatsächlich zu beobachten. Schätzungen aus den USA zufolge beläuft sich der Anteil der 'charity care' am Umsatz auf ca. 4%, wobei interessanterweise keine grossen Unterschiede zwischen non-profit und profitorientierten Spitälern festgestellt wurden [Sloan (2000), S. 1160]. Allerdings gilt es hier zu unterscheiden, wo die Behandlung tatsächlich aus altruistischen Motiven stattfindet und wo eine Aufnahmepflicht für die Spitäler besteht.

<sup>13</sup>Streng mathematisch müsste in der Zielfunktion des Auftraggebers nach Einsetzen der Nutzenfunktion (6) der Term  $V(q, n)$  auftauchen. Wenn die Bewertung des Spitals jedoch altruistisch motiviert ist

$$W = B(q, n) - (1 + \alpha)(c(e, q, n)) - \alpha t - \varphi(e, q, n). \quad (7)$$

Die Höhe der Vergütung geht im Nettoeffekt negativ in die Zielfunktion des Auftraggebers ein, so dass der Auftraggeber einen möglichst niedrigen Wert von  $t$  anstrebt.<sup>14</sup> Er wird  $t$  also so niedrig wählen, dass der Nutzen des Spitals aus der Durchführung des Auftrages gerade dem Reservationsnutzen entspricht. Mit  $s$  als Reservationsnutzen des Spitals ergibt sich für die minimale Vergütung  $t = s - V(.) + \varphi(e, q, n)$ . Damit ergibt sich letztendlich für die Zielfunktion des Auftraggebers im Modell von Chalkley und Malcomson

$$W = B(q, n) - (1 + \alpha)(c(e, q, n) + \varphi(e, q, n)) + \alpha V(q, n) - \alpha s. \quad (8)$$

Eine interne Lösung vorausgesetzt, ergibt Ableiten von (8) nach den relevanten Grössen  $e, q$  und  $n$  die Bedingungen erster Ordnung für die Werte, die der Auftraggeber wählen würde, wenn er die Grössen festlegen könnte (first best-Lösung),

$$-c_e(.) - \varphi_e(.) = 0 \quad (9)$$

$$B_q(.) + \alpha V_q(.) - (1 + \alpha)(c_q(.) + \varphi_q(.)) = 0 \quad (10)$$

$$B_n(.) + \alpha V_n(.) - (1 + \alpha)(c_n(.) + \varphi_n(.)) = 0, \quad (11)$$

wobei der unten stehende Index die Ableitung nach der jeweiligen Grösse darstellt.

**4.1.3.2 Anreizverträglichkeitsbedingungen des Spitals** Im Modell von Chalkley und Malcomson herrscht asymmetrische Information hinsichtlich der vom Spital bestimmten Qualität  $q$  und den Anstrengungen  $e$ . Bezüglich der Anzahl der behandelten Patienten  $n$  nehmen Chalkley und Malcomson an, dass sie der Auftraggeber zwar nicht vertraglich

---

und ein Interesse des Spitals an dem widerspiegelt, was gesellschaftlich gewünscht wird, sollte dieses Interesse an sich nicht bestimmend dafür sein, was gesellschaftlich gewünscht ist (Chalkley/Malcomson (1998a)). So wird der Term  $V(.)$  oft aus der Zielfunktion des Auftraggebers ausgeschlossen, wobei auf einen entsprechenden Artikel von Hammond verwiesen wird [Hammond (1987)].

<sup>14</sup>Der Auftraggeber ist zwar bei einem risikoneutralen Spital grundsätzlich indifferent gegenüber der Höhe der Vergütung, da seine gesamtgesellschaftliche Zielfunktion (5) auch den Nutzen des Spitals einschliesst. Bei positiven Schattenkosten  $\alpha$  entstehen dem Auftraggeber bei der Vergütung des Spitals allerdings „Reibungsverluste“, so dass er einen möglichst tiefen Wert von  $t$  anstrebt. Lediglich wenn die Schattenkosten bei 0 liegen, ist der Auftraggeber indifferent gegenüber der Höhe der Vergütung, denn dann stellt die Vergütung lediglich ein Parameter zur Verschiebung des Nutzens der Patienten gegen den Nutzen des Spitals dar, ohne die Summe aus beiden Nutzen zu verändern. Dies ist zum Beispiel bei Ma (1994) der Fall.

festschreiben, jedoch beobachten und neben den Behandlungskosten als zusätzliches Argument für die Vergütung einsetzen kann.

Eine Allokation  $\{q^*, e^*, n^*\}$  ist für den Auftraggeber mit dem Vergütungssystem  $t = t(c, n)$  nur zu erreichen, wenn

$$(q^*, e^*, n^*) \in \arg \max_{e, q, n} t(c(\cdot), n) + V(q, n) - \varphi(e, q, n) \quad (12)$$

erfüllt ist.

Wenn  $q^*, e^*, n^*$  im Maximierungsprogramm des Spitals eine interne Lösung ist, dann lassen sich die in (12) zusammengefassten Anreizverträglichkeitsbedingungen auch als Bedingungen erster Ordnung des Maximierungsprogramms des Spitals darstellen:

$$t_c(c^*(\cdot), n^*)c_e^*(\cdot) - \varphi_e^*(\cdot) = 0 \quad (13)$$

$$t_c(c^*(\cdot), n^*)c_q^*(\cdot) + V_q^*(\cdot) - \varphi_q^*(\cdot) = 0 \quad (14)$$

$$t_c(c^*(\cdot), n^*)c_n^*(\cdot) + t_n(c^*(\cdot), n^*) + V_n^*(\cdot) - \varphi_n^*(\cdot) = 0. \quad (15)$$

wobei  $c^*(\cdot) \equiv c(q^*, e^*, n^*)$  etc.

**4.1.3.3 Divergenzen der Ziele von Auftraggeber und Spital** Die Problematik des Auftraggebers im Modell von Chalkley und Malcomson besteht nun darin, mit der geeigneten Bestimmung der Vergütungsfunktion  $t(c, n)$  die bestmögliche Allokation zu erreichen. Zielkonflikte zwischen Auftraggeber und Spital bleiben dabei bestehen, wenn es nicht möglich ist, die Vergütungsfunktion  $t(c, n)$  so zu gestalten, dass die Bedingungen erster Ordnung aus dem Maximierungsproblem des Spital (13) - (15) mit den Bedingungen erster Ordnung für die first best-Lösung (9) - (11) übereinstimmen. Aus einem Vergleich der Bedingungen (13) und (9) folgt, dass ein Vergütungssystem, das mit  $t_c(\cdot) = -1 \forall c$  dem Spital die volle Kostenverantwortung überträgt (*prospektives* Vergütungssystem), anreizkompatibel mit dem effizienten Niveau der Anstrengungen ist, da die beiden Bedingungen bei  $t_c(\cdot) = -1 \forall c$  identisch sind. Ob ein Vergütungssystem mit  $t_c(\cdot) = -1 \forall c$  allerdings auch hinsichtlich der Qualität und der Anzahl der Patienten zu einer effizienten Allokation führt und damit die first best-Allokation möglich macht, hängt davon ab, wie sich die altruistische Komponente  $V(q, n)$  zu der Bewertung  $B(q, n)$  des Auftraggebers verhält. Sind  $V(q, n)$  und  $B(q, n)$  identisch, entstehen mit  $t_c(\cdot) = -1 \forall c$  auch keine Zielkonflikte bezüglich der Qualität. Wenn der Auftraggeber das Spital in diesem Fall mit  $t_n(\cdot) = 0 \forall n$  mit einem „Globalbudget“ vergütet und auf eine Einzelfallvergütung verzichtet, wählt das

Spital auch die effiziente Anzahl der Patienten. Dies folgt, da mit einem Vergütungssystem  $t_c(.) = -1$  und  $t_n(.) = 0$  die entsprechenden Bedingungen erster Ordnung (10) und (14) bzw. (11) und (15) identisch sind. Verhält sich die altruistische Komponente  $V(q, n)$  des Spitals zu der Bewertung  $B(q, n)$  dagegen so, dass mit dem System  $t_c(.) = -1 \forall c$  in einem Optimum Bedingung (14) nicht identisch mit (10) ist, bleiben Zielkonflikte bezüglich der Qualität bestehen. Damit wäre für den Auftraggeber die first best-Allokation nicht zu erreichen.

Im Modell von Chalkley und Malcomson steht der Auftraggeber vor einem typischen Problem der versteckten Handlung: Mit  $t_c(.)$  und  $t_n(.)$  stehen dem Auftraggeber nur zwei Instrumente zur Verfügung, dem Spital bei den drei Grössen  $e, q$  und  $n$  die richtigen Anreize zu setzen [Chalkley/Malcomson (1998a)]. Wenn das Spital in einer oder mehreren Grössen nicht „auf der Linie“ des Auftraggebers ist, kann der Auftraggeber ein Abweichen von der first best-Allokation nicht verhindern.

Die altruistische Komponente  $V(q, n)$  in der Zielfunktion (6) sorgt dafür, dass das Spital ein eigenes Interesse an einem gewissen Niveau der Behandlungsqualität hat. Je näher dieses Eigeninteresse an den Präferenzen des Auftraggebers liegt, umso geringer wird das Spannungsfeld zwischen der Bereitstellung der Behandlungsqualität und den Anstrengungen zur Effizienzverbesserung, d.h. je grösser das Interesse des Spitals an der Qualität, je mehr Kostenverantwortung kann der Auftraggeber dem Spital übertragen. Bei identischen Bewertungen wird der Trade-Off zwischen Qualität und Effizienz sogar ganz aufgehoben und eine vollständige Kostenverantwortung des Spitals ist optimal.

Für den Auftraggeber ist die altruistische Komponente also von entscheidender Bedeutung bei der Bestimmung des optimalen Vertrages, nicht nur hinsichtlich ihrer Existenz, sondern auch hinsichtlich ihrer Ausprägung. In ihrem Modell legen Chalkley und Malcomson die Ziele und das Verhalten des Spitals fest, indem sie die Zielfunktion (6) des Spitals und insbesondere die altruistische Komponente postulieren. Angesichts ihrer Bedeutung wäre es vom Standpunkt des Auftraggebers jedoch von Interesse, etwas mehr über die altruistische Komponente zu erfahren.

## 4.2 Begründung der Zielfunktion des Spitals

Chalkey und Malcomson begründen die Existenz der altruistischen Komponente in der Zielfunktion des Spitals relativ knapp mit einem „intrinsischen“ Interesse von non-profit Organisationen an Qualität und Quantität. Ausserdem, so die Autoren, beeinflusse die Entscheidungsfreiheit des Arztes die Entscheidungen (bei der Allokation der zur Behandlung eingesetzten Ressourcen) eventuell zugunsten des Patienten.

Mit ihrem Modell untersuchen Chalkley und Malcomson die Implikationen von verschiedenen starken Ausprägungen der altruistischen Motivation des Spitals für die optimale Vergütung. Für diese Fragestellung ist diese Begründung der reinen Existenz der altruistischen Komponente ausreichend, zumal Chalkley und Malcomson auch die Möglichkeit einer nicht-existenten altruistischen Komponente explizit betrachten. Für Aussagen darüber, welche Ausprägung des Altruismus zu *erwarten* wäre, reichen sie jedoch nicht aus. Da die Ausprägung des Altruismus jedoch entscheidend für die Ausgestaltung des optimalen Vertrages ist, wäre es im Modell von Chalkley und Malcomson aus Sicht des Auftraggebers wünschenswert, wenn er über genauere Informationen bezüglich der altruistischen Einstellung des Spitals verfügen könnte.<sup>15</sup> Gleiches gilt natürlich auch für die anderen Komponenten der Zielfunktion des Spitals und darüber hinaus auch für die allgemeine funktionelle Form derselben.

#### **4.2.1 Verhalten des Spitals als Ergebnis der Präferenzen mehrerer Entscheidungsträger und des spitalinternen Entscheidungsmechanismus**

Im Vergleich zu Chalkley und Malcomson liefern Zweifel und Breyer eine ausführlichere Begründung der Zielfunktion in ihrem Modell [Zweifel/Breyer (1997), Kapitel 9]. Die Autoren erläutern die Präferenzen von Besitzern, Ärzten, Management und Pflege als wichtigste Gruppen von Entscheidungsträgern im Spital hinsichtlich verschiedener Größen wie finanzieller Überschuss, Qualität oder Arbeitsbelastung und nehmen an, dass sich die Präferenzen in den Vorzeichen der entsprechenden Ableitungen der Zielfunktion des Spitals wiederfinden. Sie betonen, dass die Zielfunktion des Spitals ein Kompromiss zwischen den Präferenzen der verschiedenen im Spital tätigen Gruppen darstellt und zur Beschreibung des Verhaltens des Spitals Annahmen an den spitalinternen Entscheidungsprozess getroffen werden müssen.<sup>16</sup> Allerdings nehmen sie keinen konkreten Entscheidungsprozess an und leiten die Konsequenzen dieser Annahme ab, sondern gehen von einem nicht näher erläuterten Entscheidungsprozess aus, der dazu führt, dass die Präferenzen der im Spital tätigen Personen in der Zielfunktion so zur Geltung kommen, wie gerade beschrieben. Insofern halten es Zweifel und Breyer wie Newhouse, der ebenfalls der verschiedenen Präferenzen der Entscheidungsträger im Spital gewahr wird, den Entscheidungsprozess aber nicht explizit betrachtet [Newhouse (1970)]: „*We assume that some final solution is obtai-*

---

<sup>15</sup>Jack (2005) behandelt in seinem Modell den Altruismus des Spitals als exogene Grösse, deren Ausprägung private Information des Spitals ist. Für den Auftraggeber besteht im Modell von Jack also eine Problematik der versteckten Information, da das Spital schon vor Abschluss des Vertrages über diese Information verfügt.

<sup>16</sup>Der Begriff Entscheidungsprozess umfasst hier neben der spitalinternen Organisation und Hierarchie auch die Auswirkungen eventueller spitalinterner Informationsasymmetrien.

*ned among the tastes of the administrator, the trustees, and the medical staff, so that we can speak of the tastes of the hospitals decision maker.*“ Konsequenter Weise belassen es Zweifel und Breyer bei einer sehr allgemeinen Formulierung der Zielfunktion des Spitals und verzichten auf eine Annahme einer speziellen funktionalen Form.

Tatsächlich dürfte sich das Verhalten des Spitals aus dem Verhalten der einzelnen im Spital tätigen Personen ergeben, die gemäss ihrer idiosynkratischen Präferenzen ihre eigenen Ziele verfolgen. Die Art und Weise und das Ausmass, wie sich die Präferenzen eines Entscheidungsträgers im Verhalten des Spitals bemerkbar machen, hängen von den Möglichkeiten der Einflussnahme ab, welche die Position des Entscheidungsträgers in der organisatorischen Struktur des Spitals ermöglicht. Bei seinen Entscheidungen wird er zudem von Anreizen beeinflusst, die von Entscheidungsträgern auf höherer Hierarchieebene definiert werden oder aus den organisatorischen Strukturen folgen. Das Verhalten des Spitals resultiert also aus den Präferenzen der im Spital tätigen Personen *und* aus den spitalinternen organisatorischen Gegebenheiten, die den Entscheidungsprozess definieren. In diesem Sinne spiegelt eine postulierte Zielfunktion des Spitals wie (6) eine implizite Annahme einer bestimmten organisatorischen Struktur und eines bestimmten spitalinternen Entscheidungsprozesses wieder, der die Präferenzen der Entscheidungsträger in der postulierten gemeinsamen Zielfunktion vereinigt.

Ist man nun aber am umgekehrten Weg interessiert und stellt die Frage, *wie* die Präferenzen der Entscheidungsträger ihren Weg in die Zielfunktion des Spitals finden, ist die Modellierung eines bestimmten Entscheidungsprozesses notwendig. Bei diesem Ansatz wird die Zielfunktion also nicht als das Ergebnis eines *unbekannten* Prozesses *angenommen*, sondern als das Ergebnis eines *angenommenen* internen Entscheidungsprozesses, der die Präferenzen der einzelnen Entscheidungsträger zusammenführt, *hergeleitet*.<sup>17</sup>

Wenn der spitalinterne Entscheidungsprozess massgeblich die Zielfunktion bestimmt und aus unterschiedlichen Prozessen unterschiedliche Zielfunktionen resultieren, dann ist dies natürlich von Relevanz für einen Auftraggeber, der mittels Anreizstrukturen in der Spitalvergütung die bestmögliche Allokation erreichen will. Denn welchen Sinn haben Anreize zur Beeinflussung des Spitals hinsichtlich einer bestimmten Grösse, wenn es die spitalin-

---

<sup>17</sup>Diese Arbeit konzentriert sich auf die Modellierung der Zielfunktion des Spitals. Dieselbe Diskussion könnte man grundsätzlich auch im Hinblick auf die Zielfunktion des Auftraggebers führen, wenn der Auftraggeber ebenfalls eine Organisation aus verschiedenen Entscheidungsträgern ist (wie z.B. eine Versicherung). Geht man allerdings vom Auftraggeber als einem wohlwollenden Sachwalter aus, der die Maximierung der gesellschaftlichen Wohlfahrt als Ziel verfolgt, kann man auf eine in der Theorie bereits diskutierte Wohlfahrtsfunktion wie die utilitaristische Wohlfahrtsfunktion verweisen [vgl. Breyer (2005), Kapitel 2].

terne Organisation gar nicht zulässt, dass die Anreize bis zu der im Spital für diese Grösse verantwortlichen Person dringen?

#### 4.2.2 Auswirkung des internen Entscheidungsmechanismus am Beispiel der Zielfunktion des Spitals im Modell von Chalkley und Malcomson

Hierzu sei ein einfaches Beispiel genannt: Angenommen im Modell von Chalkley sind Management und Arzt die relevanten Entscheidungsträger im Spital. Der Arzt sei für die Behandlungsqualität  $q$  zuständig. Das Management sei das Führungsorgan im Spital und kommuniziere als solches mit dem Auftraggeber. Spitalintern sei es gegenüber dem Arzt in finanziellen Fragen weisungsbefugt, d.h. es kann vom Arzt die Einhaltung einer Budgetvorgabe  $c^d$  verlangen. Zudem leiste es die Anstrengung zur Verbesserung der Kosteneffizienz  $e$ , wobei wir zur Vereinfachung annehmen, die Anstrengungen  $e$  und die Anzahl der Patienten  $n$  seien exogen gegeben. Die Zielfunktion des Managements sei mit

$$G = V^m(q, n) + t - \varphi(e) - t^d \quad (16)$$

gegeben, wobei  $V^m(q, n)$  die altruistische Bewertung des Managements ist, und  $t^d$  die Vergütung des Arztes darstellt, die das Management leisten muss. Die Zielfunktion des Arztes laute

$$U = V^d(q, n) + t^d, \quad (17)$$

wobei  $V^d(q, n)$  die altruistische Bewertung des Arztes ist.

Angenommen es gelte  $\frac{\partial V^d(q, n)}{\partial q} > \frac{\partial V^m(q, n)}{\partial q} \geq 0$  sowie  $V(q, n) \equiv V^m(q, n) + V^d(q, n)$ . Dann stellt die von Chalkley und Malcomson postulierte gemeinsame Zielfunktion (6) das Ergebnis eines spitalinternen Mechanismus dar, der in der Aufsummierung der Nutzen von Arzt und Management mündet, da  $S(.) = G(.) + U(.)$ .<sup>18</sup>

Dies ist zum Beispiel bei folgendem Entscheidungsprozess der Fall: Das Management definiert die interne Budgetierung  $c^d$  für den Arzt sowie dessen Vergütung  $t^d$ , muss aber dem Arzt mindestens seinen Reservationsnutzen in Höhe von  $u$  ermöglichen. Der Arzt dagegen muss die Behandlungsqualität so wählen, dass  $c(q, e, n) = c^d$  erfüllt ist, d.h. die Qualität ist bei gegebener Anstrengung  $e$  und Anzahl der Patienten  $n$  eine Funktion der

---

<sup>18</sup>Der mit dieser gemeinsamen Zielfunktion unterstellte interne Entscheidungsmechanismus führt zwischen Arzt und Management zu einer Pareto-effizienten Allokation, bei gegebenem Vertrag des Auftraggebers.



Budgetierung:  $q \equiv q(c^d)$ . Wir nehmen an, dass ein Mindestniveau der Behandlungsqualität existiert, welches nicht unterschritten werden kann, so dass das Budget mindestens die dafür notwendige Höhe haben muss.

Da die Vergütung  $t^d$  den Nutzen des Managements schmälert, wird das Management die Vergütung so wählen, dass der Arzt gerade seinen Reservationsnutzen von  $U = u$  erhält. Damit ergibt sich durch Einsetzen von (17) in (16) für die Zielfunktion des Managements

$$G = V^m(q(c^d), n) + V^d(q(c^d), n) + t - \varphi(e) - u. \quad (18)$$

Mit der Normalisierung  $u = 0$  resultiert für (18) die Zielfunktion des Spitals (6) im Modell von Chalkley und Malcomson. Aus Sicht des Auftraggebers kommen die Präferenzen des Arztes und des Managements hinsichtlich der Behandlungsqualität in der Zielfunktion des Spitals voll zur Geltung, so dass er diese in seinem Vergütungssystem berücksichtigen kann.<sup>19</sup> Angenommen, der Auftraggeber vergüte das Spital mit einem linearen Vergütungssystem

$$t = k - rc, \quad (19)$$

mit  $k$  als eine Pauschale und  $r \in [0, 1]$  als Grad, mit dem das Spital an den Behandlungskosten beteiligt wird. Dann wählt das Management die Budgetierung  $c^d$  so hoch, dass

$$\left( \frac{\partial V^m(q, n)}{\partial q} + \frac{\partial V^d(q, n)}{\partial q} \right) \frac{\partial q(c^d)}{\partial c^d} - r = 0 \quad (20)$$

erfüllt ist (eine interne Lösung bei gegebenen  $r > 0$  vorausgesetzt), d.h. der Auftraggeber kann mit einer entsprechenden Wahl des Beteiligungsparameters  $r$  ein bestimmtes Qualitätsniveau erwirken, welches über dem Mindestniveau liegt.

Betrachten wir nun einen anderen, leicht abgeänderten Entscheidungsprozess: Nehmen wir an, die Vergütung des Arztes sei mit  $t^d = \bar{t}^d$  exogen gegeben und hoch genug, um dem Arzt auch beim Mindestniveau der Qualität den Reservationsnutzen zu ermöglichen. In diesem Fall gelangen die Präferenzen des Arztes nicht in die Zielfunktion des Spitals. Für das Management macht es keinen Sinn, die Präferenzen des Arztes zu beachten, da die exogene Vergütung des Arztes dessen Teilnahme garantiert. Mit dem Vergütungssystem (19) lautet die Zielfunktion des Managements also

---

<sup>19</sup>Dies setzt natürlich voraus, dass der Auftraggeber über die Präferenzen des Arztes und des Managements informiert ist.

$$U = V^m(q(c), n) + k - rc - \varphi(e) - \bar{t}^d. \quad (21)$$

Bei gegebenem  $r$  wählt das Spital aus Sicht des Auftraggebers die Budgetierung und damit die Qualität derart, dass

$$\frac{\partial V^m(q, n)}{\partial q} \frac{\partial q(c^d)}{\partial c^d} - r = 0 \quad (22)$$

erfüllt ist. Ein Vergleich von (22) mit (20) zeigt, dass wegen  $\frac{\partial V^d(q, n)}{\partial q} > \frac{\partial V^m(q, n)}{\partial q} \geq 0$  beim ersten Mechanismus bei gegebener Kostenbeteiligung  $r$  eine höhere Budgetierung und damit eine höhere Behandlungsqualität resultiert als beim zweiten Mechanismus. Besonders deutlich wird dies, wenn das Management indifferent gegenüber der Behandlung ist und  $V^m(q, n) \equiv 0$  gilt. Beim ersten Mechanismus sorgen aus Sicht des Auftraggebers immer noch die Präferenzen des Arztes für ein Eigeninteresse des Spitals an der Behandlungsqualität, so dass der Auftraggeber an den Präferenzen des Arztes ansetzen kann. Beim zweiten Mechanismus schlagen die Präferenzen dagegen nicht durch und der Ausdruck (22) ist immer negativ. Das Management wählt die Budgetierung also immer gerade so hoch, dass das Mindestniveau der Qualität möglich wird. Die Bemühungen des Auftraggebers, bei den Präferenzen des Arztes anzusetzen, laufen hier also ins Leere.

Dieses einfache Beispiel zeigt, welchen Einfluss der spitalinterne Entscheidungsprozess auf das Verhalten des Spitals haben kann, so dass es für den Auftraggeber durchaus von Bedeutung ist, über die spitalinternen Gegebenheiten informiert zu sein. Wir fassen zusammen:

**Folgerung 4.1** *Die Zielfunktion des Spitals ist ein Ergebnis eines spitalinternen Entscheidungsprozesses, der die Präferenzen der spitalinternen Entscheidungsträger zusammenführt. Zur (bewussten) Ausgestaltung eines optimalen Vergütungs- und Anreizsystems muss der Auftraggeber also über die Präferenzen und den implementierten Entscheidungsprozess informiert sein.*

## 5 Das Prinzipal-Agent-Modell mit doppeltem Unterstellungsverhältnis

Zur Analyse der Auswirkung des spitalinternen Entscheidungsprozesses auf die Zielsetzung des Spitals und auf das optimale Anreizsystem des Auftraggebers wird in dieser Arbeit ein einfaches PA-Modell durch eine explizite Modellierung des spitalinternen Entscheidungsprozesses als ein nicht-kooperatives Spiel zwischen Chefarzt und Spitalleitung bzw. Management erweitert. In diesem Kapitel wird das erweiterte PA-Modell vorgestellt und die Annahmen des Modells erläutert.

### 5.1 Überblick

Das Modell bildet eine Situation ab, in der ein Auftraggeber ein Spital mit der Behandlung von Patienten einer bestimmten Population beauftragt. Die Patienten haben dieselbe Krankheitsätiologie, unterscheiden sich aber in der Schwere der Erkrankung. Der Auftraggeber vergütet das Spital für jeden behandelten Patienten einzeln. Spitalintern werden die zur Durchführung des Auftrages relevanten Entscheidungen von zwei Entscheidungsträgern getroffen: Über Aufnahme, Diagnose und Behandlung eines Patienten entscheidet der Arzt, das Spitalmanagement entscheidet dagegen über die Annahme oder Ablehnung des Vertragsangebotes des Auftraggebers. Ausserdem fällt die Kommunikation mit dem Auftraggeber ausschliesslich in den Tätigkeitsbereich des Managements, zwischen Arzt und Auftraggeber besteht keine direkte Beziehung.

Bei der Behandlung eines Patienten muss der Arzt eine Budgetvorgabe beachten, die für jede Fallschwere definiert wird. Ausserdem erhält der Arzt bei der Behandlung eines Patienten zusätzliche freie Mittel vom Management, die er zur Bewirtschaftung seiner Abteilung einsetzen kann. Für die Bestimmung der Fallbudgetierung und der freien Mittel werden zwei Entscheidungsmechanismen betrachtet. Der erste Mechanismus besteht in einem Prinzipal-Agent-Verhältnis zwischen Management und Arzt, wobei das Management als Prinzipal des Arztes die beiden Parameter festlegt. Beim zweiten Mechanismus wird dagegen angenommen, die beiden Parameter werden im Rahmen einer Verhandlung zwischen Management und Arzt festgelegt. Zur Verdeutlichung stellt Abbildung 1 die beiden Versionen des Modells gegenüber.

### 5.2 Spitalinterne Entscheidungsträger Arzt und Management

In dem hier entwickelten Modell wird das Spital als eine Organisation von mehreren Entscheidungsträgern bzw. Gruppen von Entscheidungsträgern betrachtet, die im Spital verschiedene Aufgaben erfüllen und ihre Entscheidungen fällen. Allerdings werden für die

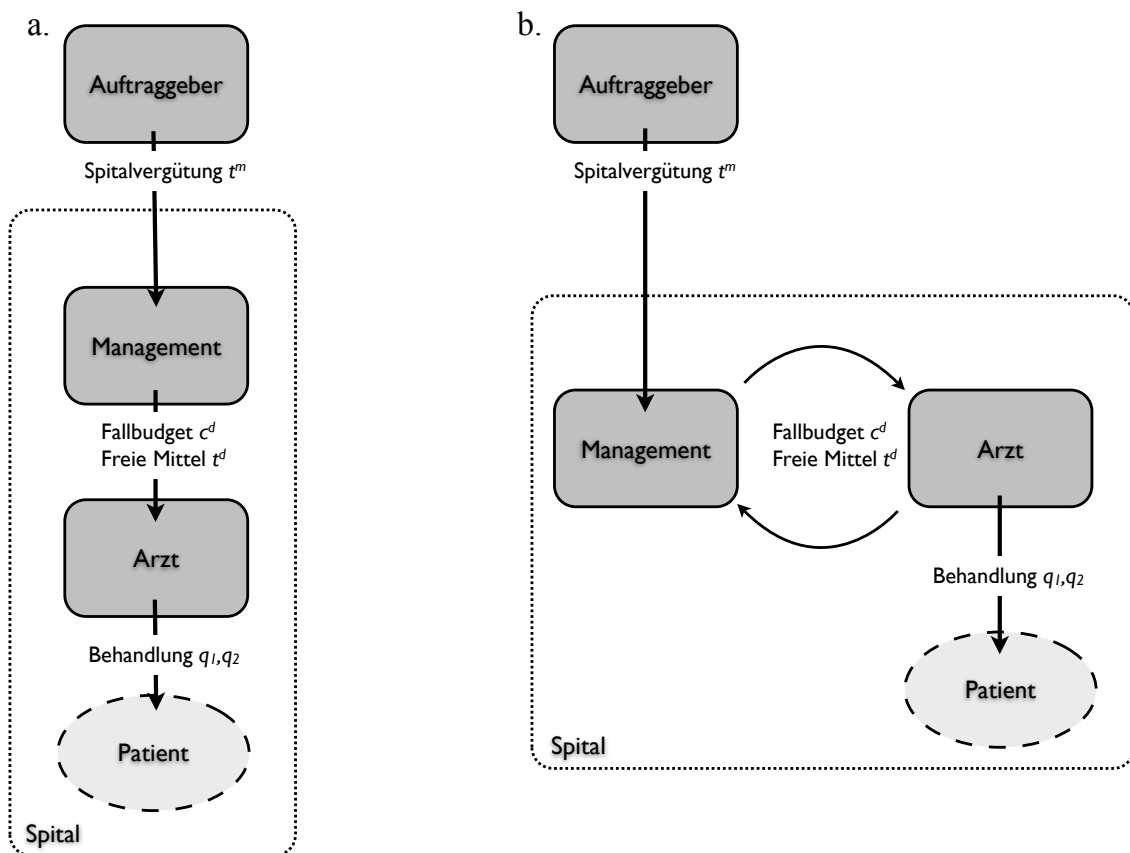


Abbildung 1: Das Modell mit a. spitalinternem Prinzipal-Agent-Verhältnis zwischen Management und Arzt und b. Verhandlungslösung zwischen Management und Arzt.

Modellierung nur die beiden für das Verhalten des Spitals wichtigsten Gruppen in die Modellierung einbezogen, namentlich das Spitalmanagement und die Ärzteschaft, wobei in beiden Bereichen jeweils die höheren Hierarchieebenen als besonders einflussreich betrachtet werden. Das Management legt die grundsätzliche *strategische Ausrichtung* des Spitals fest und ihm obliegt die *unternehmerische Führung* des Spitals. Die Ärzteschaft bestimmt dagegen grundsätzlich die *Wahl, Planung, Durchführung und Veranlassung von Diagnose- und Therapiemassnahmen*<sup>20</sup>, wobei die grundsätzlichen Entscheidungen von den leitenden Ärzten, allen voran den Chefärzten, getroffen werden [vgl. Busch (2011)].

### 5.3 Patienten und Schwere der Erkrankung

Es sei angenommen, der Auftraggeber delegiere die Behandlung einer Population von Patienten mit einer bestimmten Krankheitsätiologie<sup>21</sup> an das Spital. Die Patienten unterscheiden sich jedoch in der Schwere der Erkrankung. Die Schwere der Erkrankung oder die Fallschwere eines Patienten ist umso höher, je höher der gesellschaftliche Wohlfahrtsverlust durch Eintritt der Erkrankung ist.<sup>22</sup> Im Folgenden wird die Fallschwere mit  $\theta$  dargestellt, wobei  $\theta \in \Theta$  und  $\Theta = [\underline{\theta}, \bar{\theta}] \subset \mathbb{R}^+$ .

Die Fallschwere folgt der Verteilung  $F(\theta)$  mit der Dichte  $f(\theta)$  auf dem Intervall  $\Theta$ . Die Verteilung ist als Allgemeinwissen dem Auftraggeber, dem Management und dem Arzt bekannt.  $\theta$  lässt sich, ähnlich dem Kostengewicht in einem DRG-Vergütungssystem, als ein Gewichtungssparameter ohne eigene Masseinheit auffassen.

#### 5.3.1 Nachfrage

Die Nachfrage wird zur Vereinfachung als exogen angenommen. Sie liegt bei der betrachteten Population bei  $n$  Patienten. Sie ist dem Auftraggeber, dem Management und dem Arzt bekannt.

### 5.4 Feststellung der Fallschwere und Behandlung durch den Arzt

Wir nehmen an, der Arzt kann sich zum Zeitpunkt der Aufnahme mit der Erstellung einer ersten Diagnose ein genaues Bild von der Fallschwere des Patienten machen, bevor die endgültige Entscheidung zur Aufnahme getroffen wird und nennenswerte Ressourcen

<sup>20</sup>Damit verursachen sie direkt oder indirekt über 70% - 80% der Spitalkosten [Benz (2007)].

<sup>21</sup>Beispielsweise sei eine Diagnosefamilie wie das ICD-10 Unterkapitel I21 „Akuter Myokardinfarkt“ genannt.

<sup>22</sup>Der Begriff „Erkrankung“ schliesst Unfälle mit ein.

zur weiteren Diagnose und Behandlung verbraucht werden [vgl. Mougeot (2009)].

Wird der Patient im Spital aufgenommen, entscheidet der Arzt über weitere Diagnosemassnahmen und über die durchzuführende Behandlung. Wir nehmen an, die Krankheitsätiologie der Patienten sei recht häufig und die Therapie bzw. Behandlungsmethode sei grundsätzlich als medizinischer Standard gegeben (zum Beispiel Bypass-Implantation bei Herzinfarkt). Weiter nehmen wir an, die Behandlungsmethode bestehe aus einem Portfolio an medizinischen Leistungen, über deren Intensität der Arzt als Experte von Fall zu Fall entscheidet. In der Realität bestehen Diagnose und Behandlung im Spital aus einer Vielzahl von Massnahmen, deren Zusammensetzung und Intensitäten vom Arzt gewählt werden. Um die Analyse aber überschaubar zu halten und gleichzeitig die Multidimensionalität der Behandlung anzudeuten, nehmen wir an, die Behandlung eines Patienten bestehe im Modell aus zwei Leistungen (Leistung 1 und Leistung 2), von denen der Arzt die Mengen  $q_1$  und  $q_2$  zur Behandlung wählt.

## 5.5 Behandlungskosten

Der Preis für eine Einheit der Leistung 1 bzw. Leistung 2 beläuft sich für das Spital auf  $p_1$  bzw.  $p_2$ . Für die Kosten der Behandlung eines Patienten nehmen wir

$$c = p_1 q_1 + p_2 q_2 \tag{23}$$

an. Gegen diese Kostenfunktion mag eingewendet werden, dass sie linear in den eingesetzten Mengen ist. Zudem wird Linearität in der Anzahl der Patienten unterstellt. Auf den ersten Blick scheinen beide Einwände berechtigt zu sein, auch im Hinblick auf den grossen Anteil an (mittelfristig) fixen Kosten, die in einem Spital anfallen. Um diese Effekte abzubilden, könnte man die Preise von den Mengen und der Anzahl der behandelten Patienten abhängig machen. Da jedoch nicht klar ist, welche Vorzeichen die jeweiligen Ableitungen haben sollten und da dies die Analyse zu stark verkomplizieren würde, wird von den Preisen  $p_1$  und  $p_2$  angenommen, sie seien von der Menge und der Anzahl der Patienten unabhängig. Zudem wird zunächst angenommen, die Preise seien exogen gegeben und vom Spital in keiner Weise zu beeinflussen.

Die Preise  $p_1$  und  $p_2$  für die Einheiten  $q_1$  und  $q_2$  sind nicht nur als Marktpreise zu verstehen. Sie enthalten etwa auch Gemeinkostenzuschläge wie den spitalinternen Verwaltungsaufwand.<sup>23</sup>

---

<sup>23</sup>Aus betriebswirtschaftlicher Sicht lassen sich Preise als Verrechnungspreise der innerbetrieblichen Leistungsverrechnung zu Vollkosten interpretieren.

Im Zusammenhang mit den Kosten soll angemerkt werden, dass diese a priori unabhängig von der Fallschwere sind. Die Behandlung eines Patienten ist nicht deshalb teuer, weil er eine schwere Erkrankung hat, sondern weil die Menge der eingesetzten Mittel hoch ist [vgl. Ellis (1998)].

## 5.6 Vertrag zwischen Auftraggeber und Spital sowie Vergütung des Spitals

Für die Behandlung der  $n$  Patienten bietet der Auftraggeber dem Management einen Vertrag in Form eines take-it-or-leave-it Angebotes an. Der Vertrag definiert die Vergütung  $t^m(\theta)$  des Spitals für die Behandlung eines Patienten.

In den Kapiteln 6 und 7 konzentrieren wir uns auf ein lineares Vergütungssystem, welches die Vergütung des Spitals in einen linearen Zusammenhang mit den realisierten Behandlungskosten pro Patient bringt (cost-sharing). Bei einem Patienten der Fallschwere  $\theta$  hat es die Form

$$t^m(\theta) = k(\theta) - r(\theta)c, \quad (24)$$

wobei  $k(\theta)$  eine von den Kosten unabhängige Pauschale und  $c$  die realisierten Behandlungskosten sind und  $r(\theta)$  den Anteil der Kosten darstellt, den das Spital zu tragen hat.

Mit dem linearen System (24) bedient sich der Auftraggeber des bereits in Abschnitt 4.1.3 erwähnten Instrumentes der „Kostenverantwortung“ zur Bildung von Anreizen für das Spital. Der Auftraggeber überträgt dem Spital Kostenverantwortung, indem er eine Vergütung des Spitals definiert, die mit den realisierten Kosten ceteris paribus sinkt. Mit der Übertragung von Kostenverantwortung an das Spital generiert er somit Anreize beim Spital, die Kosten zu senken. Die Anreize sind umso stärker, je grösser die Beteiligung  $r$  ist. Bei  $r = 1$  trägt das Spital die gesamte Kostenverantwortung, da die marginalen Kosten eins zu eins im marginalen Gewinn aufgehen. In diesem Fall spricht man von einer *prospektiven* Vergütung. Bleibt die Kostenverantwortung mit  $r = 0$  vollständig beim Auftraggeber, ist die Vergütung *retrospektiv*. Ein *gemischtes* System überträgt mit  $0 < r < 1$  die Kostenverantwortung teilweise an das Spital.

Das System (24) verlangt, dass der Auftraggeber die Kosten beobachten kann. Die Beobachtbarkeit der Behandlungskosten ist eine gängige Annahme in der allgemeinen PA-Literatur, aber auch in der theoretischen Literatur zur Spitalregulierung. Diese Annahme

ist auch gut mit der Realität in Einklang zu bringen. So müssen z.B. im Kanton Zürich Spitäler mit kantonalem Versorgungsauftrag jährlich bei den kantonalen Behörden die Kosten für jeden stationären Behandlungsfall des abgelaufenen Geschäftsjahres einreichen.

Grundsätzlich könnte der Auftraggeber anstatt des linearen Systems zur Vergütung des Spitals auch ein System anwenden, das dem Spital ein Fallbudget vorschreibt und zur Vergütung eine von den Kosten unabhängige Pauschale definiert. Allerdings ist im Hinblick auf die Umsetzung in der Realität ein lineares Anreizsystem einfacher zu handhaben, da dieses System auf Anreizen beruht und Sanktionsmechanismen wegfallen, die bei einem System mit Budgetvorgaben notwendig sind.<sup>24</sup> Da der Auftraggeber zumindest bei der Problematik der versteckten Handlung, wie sie in Kapitel 6 und 7 thematisiert wird, mit den Anreizen des linearen Systems dieselbe Allokation erreichen kann wie unter dem Budgetierungssystem, wird er das lineare Anreizsystem vorziehen.

Dagegen ist die optimale Allokation, welche mit einem Budgetierungssystem zu erreichen ist, bei einer Problematik der versteckten Information, wie sie in Kapitel 8 betrachtet wird, nur unter bestimmten Voraussetzungen mit einem linearen System zu implementieren. Deshalb wird in Kapitel 8 angenommen, der Auftraggeber vergüte das Spital in Abhängigkeit der Fallschwere und verlange die Erfüllung einer Budgetvorgabe. Diese Vergütung wird in Kapitel 8 näher erläutert.

## 5.7 Spitalinterne Fallkostenbudgetierung und frei verfügbare Mittel für den Arzt

Bei der Behandlung eines Patienten der Fallschwere  $\theta$  muss der Arzt das spitalinterne Kostenbudget  $c^d(\theta)$  einhalten, d.h. er wird bei der Wahl der Mengen  $q_1$  und  $q_2$  durch die Budgetrestriktion

$$p_1 q_1 + p_2 q_2 \leq c^d(\theta) \tag{25}$$

beschränkt. Grundsätzlich ist es möglich, dass der Arzt bei der Behandlung unter dem Budget bleibt. Wir nehmen allerdings an, dass der Arzt die Mittel vollständig ausschöpft, so dass die Bedingung (25) immer bindet.<sup>25</sup>

---

<sup>24</sup>Solche linearen Vergütungssysteme haben neben ihrer Einfachheit die Eigenschaft, dass sie relativ robust gegenüber der Annahme von unsicheren Zusammenhängen zwischen den Kosten und den Aktionen des Spitals sind [vgl. dazu z.B. Laffont/Tirole (1993), Seite 72].

<sup>25</sup>Wir gehen davon aus, das Management verfüge über geeignete Sanktionsmöglichkeiten, um den Arzt



Ausserdem stellt das Management dem Arzt pro behandeltem Patient der Fallschwere  $\theta$  zusätzlich freie Mittel in Höhe von  $t^d(\theta)$  zur Verfügung. Mit den freien Mittel kann der Arzt zum Beispiel Forschungsprojekte finanzieren oder seine Abteilung bewirtschaften bzw. materiell besser ausstatten.

Die Bestimmung des Fallbudgets und der freien Mittel hängt vom jeweils angenommenen spitalinternen Entscheidungsmechanismus ab. Besteht zwischen Management und Arzt ein Prinzipal-Agent-Verhältnis, werden die Budgetierung und die freien Mittel vom Management bestimmt. Werden dagegen die beiden Parameter zwischen Management und Arzt verhandelt, übt auch der Arzt entsprechend seiner Verhandlungsmacht Einfluss bei deren Bestimmung aus.

## 5.8 Präferenzen und Zielfunktionen

Im Modell müssen Annahmen an die Zielfunktionen von Auftraggeber, Arzt und Management getroffen werden. Grundsätzlich wird angenommen, der Nutzen des Auftraggebers sowie des Arztes hänge von der Qualität der Behandlung ab. Da die Qualität der Behandlung allerdings ein schwieriges Konzept ist, erfolgt im nächsten Abschnitt eine kurze Erläuterung, wie im Modell mit diesem Thema umgegangen wird.

### 5.8.1 Qualität der Behandlung

Die Definition der Qualität der Behandlung im Spital ist eine schwammige Angelegenheit, obwohl die allgemeine mikroökonomische Definition der Qualität recht einfach klingt: Ökonomen definieren die Qualität eines Gutes im Allgemeinen als ein Attribut, dessen Erhöhung *ceteris paribus* die Nachfragekurve nach dem Gut nach aussen verschiebt.<sup>26</sup>

Grundsätzlich zählt das Ergebnis der Behandlung: Wiederherstellung der Funktionalität, Heilung, Schmerzfreiheit, aber auch Komplikationen. Doch wie bereits in der Einleitung erwähnt, besteht zwischen dem Ergebnis der Behandlung und der medizinischen Leistung ein unsicherer Zusammenhang, so dass zum Beispiel ein schlechtes Behandlungsergebnis nicht unbedingt durch eine schlechte Behandlungsqualität verursacht und vom Spital verantwortet werden muss.

---

davon abzuhalten, die Budgetvorgabe nicht einzuhalten. Welche Sanktionen dies genau sein könnten, spielt hier keine Rolle. Entscheidend ist, dass sie für den Arzt einen derart hohen Nutzenverlust bedeuten, dass sich dieser nicht zu einem beobachtbaren Fehlverhalten hinreissen lässt.

<sup>26</sup>Hier stellt sich bereits die Frage, wessen Nachfragekurve dabei relevant ist: Die vom Patienten oder diejenige vom Finanzier und Auftraggeber der Leistungen, der in der Regel nicht mit dem Patienten identisch ist.

Vom Spital zu verantworten ist dagegen die Art und Weise, wie es die Behandlung durchführt, so dass sich die Behandlungsqualität eines Spitals daran festmachen lässt. Die Behandlung eines Patienten im Spital ist ein sehr komplexer Prozess, der aus einer Vielzahl von Leistungen besteht. In erster Linie sind dies natürlich medizinische Leistungen wie beispielsweise Ultraschalluntersuchung, Röntgen, Laboranalysen oder Computertomographie aus dem diagnostischen Bereich und Medikation, Operation, Physiotherapie und Pflege Techniken aus dem therapeutischen Bereich. Dazu zählen jedoch auch auch nicht-medizinische Leistungen wie Verpflegung, Reinigung, Sterilisation, Administration, sowie Leistungen mit Investitionscharakter wie Fortbildung von Personal oder Investitionen in die Infrastruktur (medizinische Geräte, Immobilien etc.). Bei der Durchführung all dieser Leistungen können die verantwortlichen Personen im Spital mehr oder weniger Anstrengungen unternehmen und mehr oder weniger Sorgfalt walten lassen. Zudem sind auch die Koordination und die Auswahl der Leistungen entscheidende Faktoren.

In diesem Sinne lassen wir den Aspekt der Behandlungsqualität in die Analyse einfließen, in dem wir davon ausgehen, der jeweilige Betrachter bewerte die Kombination der zur Behandlung eines Patienten eingesetzten Mengen  $q_1$  und  $q_2$  und nicht das Ergebnis der Behandlung.

### 5.8.2 Präferenzen und Zielfunktion des Auftraggebers

Bevor die Ziele des Auftraggebers diskutiert werden können, muss erst einmal klar sein, wer der Auftraggeber eigentlich ist. Grundsätzlich in Frage kommen zum Beispiel eine staatliche Institution, die im Sinne der Gesellschaft handelt, eine Krankenversicherung, die mit den Prämien der Mitglieder die Behandlung finanziert oder ein Patient, der als Selbstzahler für seine eigene Behandlung aufkommt. Gegen eine Krankenkasse oder einen Patienten als Auftraggeber spricht allerdings die implizite Annahme des Prinzipal-Agent-Modells, dass sämtliche Kompetenzen bei der Gestaltung des Vertrages inklusive des Vergütungssystems beim Auftraggeber (Prinzipal) liegen und der Auftraggeber dem Spital (Agenten) den Vertrag als *take-it-or-leave-it* Angebot unterbreitet.

Derartige Kompetenzen haben eigentlich nur staatliche Institutionen, nicht aber Krankenkassen<sup>27</sup> und schon gar nicht einzelne Patienten. In der Schweiz zum Beispiel sind die Verträge in der Regel das Ergebnis von Verhandlungen zwischen Spital und Krankenkasse,<sup>28</sup> wobei es kaum vorkommt, dass sich ein Verhandlungspartner mit all seinen anfänglichen Forderungen durchsetzt. Von daher scheint für die Modellierung des Verhältnisses von Krankenkasse und Spital das PA-Modell nicht unbedingt der geeignete Rahmen

<sup>27</sup>Es sein denn, das Spital *gehört* der Krankenkasse.

<sup>28</sup>Hierbei werden die Verhandlungen häufig auf Verbandsebene geführt.

[vgl. Chalkley (2000), Seite 852]. Ein Ansatz, der stattdessen den Vertrag zwischen Spital und Krankenkasse als Ergebnis eines Verhandlungsprozesses betrachtet, scheint hier zielführender zu sein.<sup>29</sup>

Aus diesem Blickwinkel betrachtet bleibt als Auftraggeber also nur der Staat bzw. eine staatliche Institution. Gegen die Rolle des Staates als Auftraggeber mag eingewendet werden, dass in Ländern wie der Schweiz oder Deutschland der Staat nicht oder nur teilweise der Finanzier der Spitalleistungen ist.<sup>30</sup> Allerdings wird vom PA-Modell nicht gefordert, dass der Auftraggeber die Vergütung selbst leisten muss. Geht man davon aus, dass der Staat als „wohlwollender Sachwalter“ die Interessen der Mitglieder der Gesellschaft, welche ja über die Krankenkassenprämien das Spital vergüten und die Behandlung eines Patienten finanzieren, vertritt und mit gesetzlichen Vorgaben durchsetzt, dann spielt es an sich auch keine Rolle, ob nun eine staatliche Behörde die Zahlungen leistet oder eine prämiendifinanzierte Kasse dies tut. Im Sinne einer normativen Analyse scheint daher das PA-Modell ein adäquater Rahmen, in dem der Staat als Auftraggeber die Maximierung der gesamtgesellschaftlichen Wohlfahrt verfolgt [vgl. Chalkley (2000), Seite 852].

Wie viele andere Autoren übernehmen auch wir diese Sichtweise und gehen grundsätzlich davon aus, dass der Auftraggeber eine staatliche Institution ist. Hinsichtlich der Zielfunktion erlauben wir uns aber im Vergleich zu einer utilitaristischen Wohlfahrtsfunktion wie (5) im Sinne einer übersichtlicheren Darstellung der Analyse eine Vereinfachung: Wir nehmen an, dass der Auftraggeber die Differenz zwischen dem Nutzen aus der Behandlung eines Patienten und den finanziellen Aufwendungen, die er dafür entrichten muss, maximiert:

$$W(\theta) = \theta B(q_1, q_2) - t^m(\theta) - c, \quad (26)$$

wobei  $\theta B(q_1, q_2)$  den Nutzen darstellt, den der Auftraggeber aus der Behandlung eines Patienten der Fallschwere  $\theta$  im Spital erhält. Im Vergleich mit einer utilitaristischen Wohlfahrtsfunktion wie (5) führt die Funktion (26) in qualitativer Hinsicht hier zu denselben Ergebnissen, vereinfacht jedoch die Darstellung. In beide Funktionen geht die Vergütung bzw. der Nutzen des Spitals negativ ein, wobei dies bei der Wohlfahrtsfunktion (5) als Nettoeffekt der Schattenkosten öffentlicher Finanzierung  $\alpha > 0$  folgt.

In (26) besteht der Nutzen des Auftraggebers aus der Behandlung eines Patienten aus einer

---

<sup>29</sup>Barros/Martinez-Giralt (2006) geben eine kurze Einführung in Modellierung von Verhandlungen auf dem Sektor für Gesundheitsleistungen.

<sup>30</sup>In der Schweiz zum Beispiel tragen die Kantone nur ca. 50% der stationären Behandlungskosten, den anderen Teil übernimmt die Krankenkasse des versicherten Patienten. In Deutschland trägt der Staat nur die Investitionskosten. Die betrieblichen Kosten trägt zu 100% die Krankenkasse.

mit der Fallschwere  $\theta$  gewichteten Funktion  $B(q_1, q_1)$ , die als Bewertung des Auftraggebers der zur Behandlung eingesetzten Kombination der Mengen  $q_1$  und  $q_2$  ausgedrückt in Geldeinheiten gesehen werden kann. Bei der Bewertung scheint es plausibel anzunehmen, dass beide Leistungen bis zu einer bestimmten Sättigungsmenge einen positiven, aber abnehmenden Effekt haben. Ab der Sättigungsmenge wird angenommen, dass jede zusätzliche Einheit dem Patienten mehr schadet als nutzt [vgl. Ernst (2000), Seite 258].

Zur formalen Umsetzung dieser Annahme, nehmen wir an,  $B(q_1, q_2)$  sei eine strikt konkave Funktion mit einem eindeutigen Maximum bei dem Sättigungspunkt  $q_1^{\circ AG}, q_2^{\circ AG}$ :

$$B_{q_h}(q_1, q_2) \begin{cases} > 0 \text{ für } q_h < q_h^{\circ AG} \\ = 0 \text{ für } q_h = q_h^{\circ AG} \\ < 0 \text{ für } q_h > q_h^{\circ AG}, \end{cases} \quad (27)$$

sowie [vgl. Hoy (2001), Theorem 11.9]

$$B_{q_h q_h}(\cdot) < 0, \quad (28)$$

und

$$B_{q_1 q_1}(\cdot) B_{q_2 q_2}(\cdot) - B_{q_1 q_2}^2(\cdot) > 0, \quad (29)$$

wobei  $h = 1, 2$ ,  $B_{q_h}(q_1, q_2) = \frac{\partial B(q_1, q_2)}{\partial q_h}$ , bzw.  $B_{q_h q_h}(q_1, q_2) = \frac{\partial^2 B(q_1, q_2)}{(\partial q_h)^2}$  usw.

Bezüglich der medizinischen Interaktion der beiden Leistungen nehmen wir an, dass sich beide Leistungen gegenseitig in ihrer Wirkung nicht schmälern

$$B_{q_1 q_2}(\cdot) \geq 0. \quad (30)$$

Diese Annahme wird vor allem getätigt, um in der späteren Analyse ein eindeutiges Verhalten des Arztes zu gewährleisten. Allerdings scheint sie auch nicht unplausibel. Hier einige Beispiele:

- Diagnose und Therapie: Je mehr Aufwand für die Diagnose betrieben wird, desto zielgerichteter kann die Therapie erfolgen.
- Arbeitszeit und medizinisches Material: Die Arbeitszeit des Arztes wird produktiver, wenn er weniger Einschränkungen beim Einsatz von medizinischem Material beachten muss (zumindest solange die Menge des medizinischen Materials unterhalb der Sättigungsmenge liegt.)

- Medizinische Therapie und Aufenthaltsdauer: Die Wirksamkeit einer medizinischen Therapie wie einer Operation wird nicht geringer, wenn der Patient länger im Spital verweilt.

Zur Bewertung der Behandlung eines Patienten gewichtet der Auftraggeber den Nutzen aus der Behandlung mit der Fallschwere des behandelten Patienten. Oben wurde die Fallschwere als ein Parameter definiert, der den Wohlfahrtsverlust durch die Erkrankung eines Patienten bestimmt, wobei der Wohlfahrtsverlust mit der Fallschwere steigt. Folglich ist es plausibel anzunehmen, dass für einen Auftraggeber, der an der Maximierung der Wohlfahrt interessiert ist, die Zahlungsbereitschaft für eine Behandlung mit gegebenen Quantitäten  $q_1$  und  $q_2$  mit der Fallschwere steigt.<sup>31</sup>

### 5.8.3 Präferenzen und Zielfunktion des Arztes

Ärzten wird gerne unterstellt, sie lassen sich bei der Behandlung von der *ärztlichen Ethik*, Altruismus oder etwas leiten, das 'im besten Interesse des Patienten' ist. Es wird unterstellt, dass der Arzt Genugtuung empfindet, wenn er sein medizinisches Wissen und Können dazu einsetzt, den Patienten bestmöglich zu behandeln [vgl. McGuire (2000), Seite 522], und zwar nicht im Sinne von 'je mehr, je besser', sondern im Sinne von 'nach vorhandenem medizinischen Wissen genau richtig'.<sup>32</sup> Wenn die ärztliche Ethik die einzige Motivation und der Nutzen des Patienten das alleinige Ziel des Arztes bei der Behandlung wäre, dann würde er dem Patienten gerade so viel medizinische Leistungen zukommen lassen, bis dessen Grenznutzen null ist.<sup>33</sup>

Neben dem Gutmenschentum sind für den Arzt jedoch noch weitere Ziele und Präferenzen plausibel. Eine wichtige Rolle spielen sicherlich auch Faktoren wie *Reputation* und Anerkennung. Der Beruf des Arztes ist und bleibt prestigeträchtig, und den meisten Ärzten dürfte dieser Umstand kaum der Grund für eine Überlegung gewesen sein, eventuell doch etwas anderes als den ärztlichen Beruf zu erlernen. Ebenso wird *Macht* und Einflussnahme als nicht-altruistische Motivation von leitenden Ärzten und vor allem von Chefärzten (oder solchen, die es werden wollen) erwähnt [Benz (2007), Seite 86]. Macht zieht ein Chefarzt zum Beispiel aus der Grösse seiner Abteilung und des Beitrages seiner Abteilung zum Umsatz des Spitals [Zweifel/Breyer (1997), Seite 283] oder aus seinem Beitrag zur

---

<sup>31</sup>Mit der linearen Gewichtung entspricht die Fallschwere der Grenzrate der Substitution zwischen den finanziellen Aufwendungen und dem Nutzen aus der Behandlung.

<sup>32</sup>Hier wird auch gerne einmal der Hippokratische Eid bemüht [z.B. Mougeot (2009)]. Dieser spielt allerdings in seiner ursprünglichen Form im Berufsleben eines Arztes keine allzu grosse Rolle mehr [Benz (2007), Seite 89].

<sup>33</sup>Wobei mancherorts die Ethik eher als Restriktion bei der Erwirtschaftung des Einkommens in Betracht gezogen wird [vgl. McGuire (2000), Seite 521].

Reputation des Spitals.

Negativ wird der Arzt eine sehr hohe Arbeitsintensität bewerten. Eine Zunahme der Aufgaben bei gegebener Arbeitszeit bedeutet weniger Zeit pro Aufgabe. Die damit verbundene höhere Arbeitsgeschwindigkeit und der resultierende Stress können eine erhebliche psychische und physische Belastung für den Arzt darstellen [Kienzle (2007)].

Nicht zuletzt ist das persönliche Einkommen eine wichtige Zielgrösse für den Arzt.

Für das Modell nehmen wir an, der Arzt maximiere bei der Behandlung eines Patienten der Fallschwere  $\theta$  die Funktion

$$U(\theta) = \theta V(q_1, q_2) + t^d(\theta) + \bar{t}^d \quad (31)$$

wobei  $\theta V(q_1, q_2)$  ähnlich wie beim Auftraggeber die Bewertung der Behandlung darstellt. Wie bereits erwähnt sind  $t^d(\theta)$  die freien Mittel. Von der persönlichen Vergütung des Arztes  $\bar{t}^d$  nehmen wir an, sie sei modell-exogen definiert. Dies entspricht der Situation in kantonalen Spitälern der Schweiz. Dort ist das Einkommen der Chefärzte unabhängig von der Leistungserbringung im Grundversicherungsbereich. Im Folgenden wird das Einkommen auf null normalisiert,  $\bar{t}^d = 0$ .

Wir nehmen an, der Arzt bewerte die Behandlung an sich mit  $V(q_1, q_2)$ , gemessen in Geldeinheiten. Wie im Fall des Auftraggebers nehmen wir an,  $V(q_1, q_2)$  sei streng konkav und besitze einen Sättigungspunkt bei  $q_1^{od}, q_2^{od}$ :

$$V_{q_h}(q_1, q_2) \begin{cases} > 0 \text{ für } q_h < q_h^{od} \\ = 0 \text{ für } q_h = q_h^{od} \\ < 0 \text{ für } q_h > q_h^{od}, \end{cases}$$

sowie

$$V_{q_h q_h}(\cdot) < 0, \quad (32)$$

und

$$V_{q_1 q_1}(\cdot) V_{q_2 q_2}(\cdot) - V_{q_1 q_2}^2(\cdot) > 0. \quad (33)$$

Auch bei der ärztlichen Bewertung nehmen wir bezüglich der medizinischen Interaktion der beiden Leistungen an, dass sich beide Leistungen gegenseitig in ihrer Wirkung nicht schmälern

$$V_{q_1 q_2}(\cdot) \geq 0. \quad (34)$$

Die Bewertung umfasst grundsätzlich die oben genannten nicht-finanziellen Ziele, die ein Chefarzt mit der Behandlung bzw. Behandlungsvorgabe für die in seinem Bereich tätigen Ärzte verfolgt: Erfolgreiche Behandlung, Reputation und Macht. Aus der altruistischen Motivation heraus sowie im Interesse seiner Reputation dürfte der Arzt eine Ausweitung der Mengen  $q_1$  und  $q_2$  bis zu einem gewissen Punkt positiv bewerten, wenn dies den Behandlungserfolg wahrscheinlicher macht. Zusätzlich dürfte eine erfolgreiche Behandlung auch der Machtposition des Arztes innerhalb des Spitals zuträglich sein. Mit zunehmender Menge gelangen aber auch Faktoren vermehrt zur Geltung, die den zusätzlichen Nutzen des Arztes aus einer Ausweitung der Mengen schmälern, wenn nicht sogar negativ werden lassen. Wie bereits angesprochen überwiegen bei vielen medizinischen Leistungen ab einer gewissen Menge die unerwünschten Nebenwirkungen, so dass eine Ausweitung der Leistungen dem Behandlungserfolg abträglich ist. Daneben ist es vorstellbar, dass eine Ausweitung der Mengen für den Arzt mit einer höheren Arbeitsintensität verbunden ist, so dass der Nutzen, den der Arzt aus den eingesetzten Mengen erfährt, geschmälert wird [vgl. Chalkley/Malcomson (2000)]. Dies betrifft intuitiv vor allem Leistungen, bei denen der Einsatz der ärztlichen Arbeitskraft mit der Menge ansteigt. Für relativ geringe Mengen der Leistungen wird dieser Effekt jedoch nicht stark ausgeprägt sein, für eine hohe Behandlungsintensität ist es vorstellbar, dass sich die Arbeitsbelastung bemerkbar macht und sich negativ auf die marginale Wertung auswirkt, so dass auch beim Arzt ein Sättigungspunkt existieren wird. Insgesamt erscheint so eine konkave Funktion  $V(q_1, q_2)$  mit einem eindeutigen Maximum eine plausible Annahme.

Auch beim Arzt gehen wir davon aus, dass er die Behandlung eines schweren Falles höher bewertet als die Behandlung eines Falles mit leichter Fallschwere, da ein Patient in der Regel mehr von der Behandlung profitiert, je ernsthafter er erkrankt ist. Zudem könnte es der Arzt auch aus einer egoistischen Motivation heraus interessanter finden, schwere Fälle zu behandeln, da ihm diese wohl mehr Möglichkeiten bieten, sein Können unter Beweis zu stellen. Wie beim Auftraggeber wird dies durch eine lineare Skalierung der Funktion  $V(q_1, q_2)$  mit der Fallschwere  $\theta$  dargestellt.<sup>34</sup>

---

<sup>34</sup>Zur Vereinfachung nehmen wir an, der Arzt beziehe sich bei der Gewichtung der Fallschwere auf das selbe Bewertungssystem wie der Auftraggeber, das heisst ist der Parameter  $\theta$  in der Nutzenfunktion des Arztes (31) ist identisch mit jenem in der Zielfunktion des Auftraggebers (26).

Die Funktion (31) unterstellt Risikoneutralität des Arztes gegenüber den freien Mitteln  $t^d(\theta)$ . Diese Vereinfachung scheint vertretbar, wenn man sich vor Augen hält, dass weder die Existenz seiner Abteilung noch sein persönliches Gehalt von der Höhe von  $t^d(\theta)$  abhängen. Wenn man so will, beeinträchtigen begrenzte freie Mittel lediglich den „unternehmerischen“ Spielraum des Arztes. Zudem ist das „Portfolio“ des Arztes mit  $n$  Patienten sehr stark diversifiziert. In Abschnitt 6.5 wird die Annahme der Risikoneutralität trotzdem zugunsten der Annahme eines in den finanziellen Mitteln risikoscheuen Arztes aufgegeben, um die Auswirkung der Annahme zu analysieren.

#### 5.8.4 Präferenzen und Zielfunktion des Managements

Es erscheint sinnvoll, die Präferenzen des Spitalmanagements anhand der gängigen Einteilung in die drei Ebenen

1. Managementebene: Spitalträgerorgane
2. Managementebene: Spitalleitung (Direktion)
3. Managementebene: Fachabteilungen / Zentren

zu diskutieren [Schmidt-Rettig (2008)], wobei hier vor allem die erste und die zweite Managementebene von Bedeutung sind, da auf diesen Ebenen schwerpunktmässig die Managemententscheidungen getroffen werden, die das Verhalten des Spital prägen. Auf die dritte Ebene wird hier deshalb nicht näher eingegangen.

**5.8.4.1 Spitalträger** Beim Spitalträger handelt es sich meist um eine juristische Person. Der Spitalträger ist der Eigentümer und Betreiber des Spitals. Mit dem Betrieb eines Spitals verfolgt der Träger bestimmte Ziele, die aus gesundheitspolitischen, besonderen gesellschaftlichen Wertvorstellungen oder ökonomischen Motivationen entspringen. Die Motivation und die Ziele bestimmen grundlegend das Verhalten des Spitals, oder wie Sonnentag [Sonnentag (2008)] beschreibt: „Die Motivation des Trägers, ein Krankenhaus zu besitzen und zu betreiben, entfaltet dabei eine Strahlkraft auf alle Entscheidungen innerhalb des Krankenhausbetriebs.“

Ein wichtiger Indikator für die Motivation des Trägers ist der rechtlich-wirtschaftliche Status. In dieser Hinsicht lassen sich in der Schweiz die

- öffentlich-rechtlichen
- privatrechtlichen, subventionierten
- privatrechtlichen, selbständigen



Spitäler unterscheiden [Undritz (2006)].<sup>35</sup> Bei den Trägern öffentlich-rechtlicher Spitäler (v.a. der Kantone) stehen *gemeinwirtschaftliche Interessen* im Vordergrund, wie Sicherstellung der medizinischen Versorgung oder der Forschung und Lehre. Angesichts sich intensivierendem Rechtfertigungsdruck auf die öffentlichen Behörden für die Verwendung steuerlicher Mittel rückt zusätzlich immer weiter die *Effizienz* der Spitäler in den Fokus der öffentlichen Träger. Dagegen sind selbständige privatwirtschaftliche Spitäler im Interesse ihrer privaten Träger tätig. Diese verfolgen mit dem Betrieb eines Spitals in erster Linie eine *Gewinnerzielungsabsicht* [Sonnentag (2008)]. Bei den Trägern privatrechtlicher Spitäler, die von der öffentlichen Hand Subventionen erhalten (in Form einer Betriebsgarantie oder Defizitdeckung) und deren Ziel nicht die Gewinnmaximierung ist, ist die Motivation eher in ideellen Werten zu suchen. Die Subventionierung durch die öffentliche Hand verpflichtet diese Spitäler aber auch, ähnlich wie öffentliche Spitäler einen Versorgungsauftrag zu erfüllen.

Die Interessen des Trägers werden durch die Mitglieder der jeweiligen Trägerorgane (1. Managementebene) wahrgenommen. Die Aufgaben sind normativer und grundsätzlich strategischer Natur. Am Universitätsspital Zürich zum Beispiel ist die erste Managementebene der Spitalrat, dessen Mitglieder vom Zürcher Kantonsrat gewählt werden. Seine Aufgaben sind unter anderem die Festlegung der Unternehmensstrategie, die Ernennung der Mitglieder der Spitaldirektion (2. Managementebene) und der Chefärzte sowie die Ausübung der Aufsicht der Spitaldirektion.

**5.8.4.2 Spitaldirektion** Die Spitaldirektion ist auf der zweiten Managementebene tätig. Die Aufgaben umfassen die operative Umsetzung der strategischen Grundsatzentscheide des Trägers sowie die unternehmerische Führung des Spitals im Sinne des Trägers [Schmidt-Rettig (2008)]. Ausserdem vertritt die Spitaldirektion das Spital nach aussen, insbesondere schliesst sie Verträge mit den Kostenträgern ab. In der Regel sind in der Spitaldirektion die Ärzteschaft, die Pflege und die Verwaltung durch die entsprechenden Direktoren vertreten. Oberstes Mitglied der Spitaldirektion ist der Spitaldirektor (Geschäftsführer, CEO), der letztendlich die Entscheidungen der Spitaldirektion trifft. Der Spitaldirektor kann gleichzeitig auch der Direktor eines Bereiches (z.B. von der Verwaltung), oder aber losgelöst von den Bereichen sein.

Die Präferenzen der einzelnen Mitglieder der Spitaldirektion werden aufgrund ihrer unterschiedlichen beruflichen Herkunft recht heterogen sein. Wie sich die einzelnen Präferenzen auf die Entscheidungen der Spitaldirektion auswirken, hängt nicht zuletzt vom Durchsetzungs- und Überzeugungsvermögens der Persönlichkeiten ab, wobei die Präferenzen

---

<sup>35</sup>In anderen Ländern finden sich vergleichbare Unterscheidungen, für Deutschland zum Beispiel siehe Sonnentag (2008).

zen des Spitaldirektors natürlich die grösste Gewichtung haben. Allen Mitgliedern dürfte aber eine mehr oder weniger starke Adaption der *Ziele des Trägers* zu eigen sein, entweder aus Überzeugung oder aus dem Gedanken heraus, ein erfolgreiches Handeln im Sinne des Trägers komme der eigenen Reputation zugute. Ausserdem werden die Präferenzen auch hinsichtlich der *Reputation* und der Bedeutung des Spitals in der Spitallandschaft übereinstimmend sein, wobei sich Bedeutung und Reputation aus der Grösse des Spitals oder führenden Positionen in der Spitzenmedizin ableiten lassen [vgl. Newhouse (1970)].

Die Sicherung der finanziellen Basis zur Aufrechterhaltung des Betriebs ist eine Kernverantwortung der Spitaldirektion, auch bei öffentlichen Spitälern und solchen Spitälern, die Subventionen aus öffentlicher Hand erhalten. Zudem verringert ein besseres finanzielles Ergebnis die Abhängigkeit von öffentlichen Mitteln und damit die Einflussnahme des öffentlichen Geldgebers auf die Geschäftstätigkeit. Eine Lockerung des finanziellen Korsetts bringt mehr unternehmerische Freiheiten, da mehr Mittel für den Betrieb des Spital zur Verfügung stehen. Von daher ist der *finanzielle Gewinn* zumindest in einem gewissen Rahmen ein zentrales Anliegen der Spitaldirektion, auch wenn es sich um ein öffentliches Spital handelt.

**5.8.4.3 Zielfunktion Management** Eigentlich besteht zwischen dem Spitalträger als Eigentümer und der Spitaldirektion als oberstem operativen Entscheidungsgremium ein klares Prinzipal-Agent-Verhältnis.<sup>36</sup> Konsequenterweise sollte es auch als solches modelliert werden, um ebenso wie die Zielfunktion des Spital als Ergebnis spitalinterner Entscheidungsprozesse die Zielfunktion des Managements als Ergebnis der Entscheidungsfindung zwischen Spitalträger und Spitaldirektion zu fundieren. Trotzdem wird in dieser Arbeit davon abgesehen, da hier der Fokus auf das Zusammenspiel des Managements mit dem Arzt mit seiner stark medizinisch geprägten Sichtweise liegt.

Insofern postulieren wir mit

$$G(\theta) = t^m(\theta) - t^d(\theta) \tag{35}$$

eine Zielfunktion, die als Ergebnis aus dem Prinzipal-Agent-Verhältnis zwischen Spitalträger und Spitaldirektion vor allem ein Interesse des Managements an dem finanziellen Überschuss aus der Behandlung eines Patienten annimmt.

---

<sup>36</sup>In der Industrieökonomik ist die Trennung von Eigentum und Kontrolle und die daraus resultierende Prinzipal-Agent-Problematik ein zentraler Aspekt bei den Ansätzen zur Erklärung des Verhaltens von Firmen [vgl. Holmstrom/Tirole (1989)].

Diese Zielfunktion greift im Hinblick auf die Präferenzen des Managements gegenüber der Art und Qualität der Behandlung natürlich zu kurz. Unbestritten hat das Management auch andere Interessen an der Behandlung als finanzielle, so dass auch hier eine ähnliche Bewertungsfunktion wie die des Arztes oder des Auftraggebers durchaus ihre Berechtigung hätte. Wenn man allerdings Präferenzen des Managements gegenüber der Behandlung annimmt, muss man aber auch die Frage beantworten, wie das Management ohne medizinisches Expertenwissen die Präferenzen bilden soll.

Diese Diskussion wird hier vermieden, indem mit der Zielfunktion (35) unterstellt wird, das Management sei indifferent bei der nicht-finanziellen Bewertung der Mengen  $q_1$  und  $q_2$ . Die Annahme vereinfacht die Analyse, hat jedoch Einfluss auf die Auswirkung asymmetrischer Informationsverteilung zwischen Arzt und Management. Auf diesen Umstand wird in Abschnitt 6.3.7 näher eingegangen.

Die Risikoneutralität des Managements lässt sich mit seinen sehr diversifizierten „Aktiva“ begründen: Allein in der Abteilung des hier betrachteten Chefarztes werden  $n$  Patienten behandelt. In einem Spital hat es das Management in der Regel jedoch mit einer Vielzahl von Chefärzten und Abteilungen zu tun, so dass das Management aus einem vielfachen von  $n$  Patienten einzeln Erträge generiert.

## 5.9 Teilnahmebedingungen

### 5.9.1 Teilnahmebedingung des Arztes

Wir gehen davon aus, dass es dem Arzt möglich sei, die Fallschwere bei der Aufnahme des Patienten abzuschätzen und den Patienten ggf. abzuweisen, wenn er in der Behandlung keinen genügenden Nutzen sieht. Damit der Arzt den Patienten nicht abweist,<sup>37</sup> muss die Behandlung des Patienten für den Arzt mindestens einen Nutzen in Höhe des Reservationsnutzens  $u > 0$  erbringen. Wegen (31) muss daher gesichert sein, dass bei der Behandlung eines Patienten der Fallschwere  $\theta$  die Teilnahmebedingung des Arztes

$$\theta V(q_1, q_2) + t^d(\theta) \geq u \tag{36}$$

erfüllt ist.

Für einen im Verhältnis zur ärztlichen Wertung der Behandlung niedrigen Reservationsnutzen ist es möglich, dass die Bedingung (36) auch noch bei einem negativen Wert der freien Mittel  $t^d(\theta) < 0$  erfüllt ist. In diesem Fall ist der Arzt also grundsätzlich dazu bereit,

---

<sup>37</sup>Was die Konsequenzen der Abweisung bedeuten, wird in Kapitel 8.3 diskutiert.

einen Abfluss der verfügbaren finanziellen Mittel seiner Abteilung hinzunehmen, solange sein Nutzen aus der Behandlung gross genug ist. Als Konsequenz würden „Schulden“ der Abteilung des Chefarztes beim Management entstehen, was aus buchhalterischer Sicht eigentlich nicht möglich ist. Zwar gibt es in der Praxis Konzepte zur Lösung des Problems, jedoch nehmen wir einen Reservationsnutzen an, der hoch genug ist, dass bei bindender Bedingung (36) negative Werte für die freien Mittel  $t^d(\theta)$  nicht vorkommen.<sup>38</sup>

### 5.9.2 Teilnahmebedingungen des Managements

Beim Management müssen zwei Teilnahmebedingungen beachtet werden. Die erste bezieht sich auf die interne Interaktion mit dem Arzt. Wir nehmen an, dass die Behandlung eines Patienten mindestens einen finanziellen Gewinn in Höhe von  $a$  erbringen muss. Für die Behandlung eines Patienten der Fallschwere  $\theta$  gilt damit die *interne* Teilnahmebedingung des Managements:

$$G(\theta) \geq a. \quad (37)$$

Für den Fall, dass diese Bedingung bei einem Patienten der Fallschwere  $\theta$  nicht erfüllt ist, nehmen wir an, das Management könne dem Arzt die Behandlung von Patienten dieser Fallschwere untersagen. Da diese Bedingung nur dann zum Tragen kommt, wenn Management und Arzt die Budgetierung  $c^d(\theta)$  und die freien Mittel  $t^d(\theta)$  im Rahmen einer Verhandlung festlegen, wird auf die interne Mindestgewinnbedingung in Abschnitt 6.4 näher eingegangen.

Die zweite Teilnahmebedingung hat im Zusammenhang mit dem Vertragsangebot des Auftraggebers Bedeutung. Sie verlangt, dass dem Spital bei der Durchführung des Auftrages genügend „Spielraum“ bleibt. Konkret verlangt die Bedingung, dass die Behandlung eines Patienten für Arzt und Management mindestens einen gemeinsamen Nutzenüberschuss in Höhe von  $g > 0$  erbringen muss:

$$N\ddot{U}(\theta) = G(\theta) - a + U(\theta) - u \geq g. \quad (38)$$

Für den Fall, dass die Bedingung (38) für einen Patienten der Fallschwere  $\theta$  nicht erfüllt ist, stimmt das Management dem Vertragsangebot des Auftraggebers nicht zu bzw. verweigert dem Arzt die Zusage für die Aufnahme des Patienten.<sup>39</sup>

---

<sup>38</sup>Ein hoher Reservationsnutzen spricht für einen Chefarzt mit hoher fachlicher Reputation, da sich einem renommierten Arzt viele hochwertige Alternativen zur Arbeit im Spital bieten dürften.

<sup>39</sup>Was die Konsequenzen für das Management aus der Ablehnung bedeuten, wird wie die Konsequenzen für den Arzt in Kapitel 8.3 diskutiert.

Der Reservationsgewinn des Managements sei zur Vereinfachung ohne Beschränkung der Allgemeinheit im Weiteren auf

$$a = 0 \tag{39}$$

gesetzt.

## 5.10 Mechanismen der Entscheidungsfindung

In einem Spital liegt die Kompetenz der Ressourcenzuteilung an die klinischen Abteilungen normalerweise beim Management. Die Zuteilung wird hauptsächlich über eine finanzielle Budgetierung umgesetzt, die für den Chefarzt bindend ist [Kienzle (2007)].

Die Ärzteschaft, insbesondere der Chefarzt, ist für den Einsatz der Ressourcen in der klinischen Abteilung verantwortlich, muss dabei allerdings vor allem die Budgetrestriktion des Managements beachten. In medizinischen Fragen ist das Management dem Chefarzt allerdings nicht weisungsbefugt [Kersting (2008)], und könnte dies aufgrund der Expertise der Chefärzte auch kaum sein (Benz (2007), Seite 97). In ihrer wesentlichen Tätigkeit arbeiten die Chefärzte also weitgehend weisungsfrei und autonom [Kersting (2008)].

Mit dieser Kompetenzregelung ist es naheliegend, auch das spitalinterne Verhältnis zwischen Chefarzt und Spitalmanagement im Rahmen des Prinzipal-Agent-Modells zu behandeln. Dabei überträgt das Management (vor allem der Spitalträger) als Prinzipal die Behandlung eines Patienten an den Chefarzt als Agenten. Hierfür definiert das Management (vor allem die Spitaldirektion) eine finanzielle Vorgabe und gegebenenfalls weitere Zuwendungen an den Chefarzt und seine Abteilung. Die Durchführung des Auftrages, also die Behandlung von Patienten innerhalb der finanziellen Vorgabe, bestimmt dagegen der Arzt und ist vom Management vertraglich nicht zu regeln. Hinzu kommt, dass der Arzt relevante Information erlangt, die dem Management nicht zugänglich sind.

Im hier entwickelten Modell wird der interne PA-Mechanismus umgesetzt, in dem das Management als Prinzipal die optimale Budgetierung  $c^d(\theta)$  und die optimalen freien Mittel  $t^d(\theta)$  unter Berücksichtigung der Anreizverträglichkeitsbedingung und der Teilnahmebedingung des Arztes als Agent wählt.

Die Achillesferse bei der Verwendung des PA-Modells als Grundlage des internen Entscheidungsmechanismus ist die Annahme des PA-Modells, dass der Prinzipal (das Ma-

nagement) die Parameter des Vertrages bestimmt und dem Agenten (dem Chefarzt) als ein take-it-or-leave-it Angebot unterbreitet. Damit würde im Modell die gesamte Verhandlungsmacht bei der Bestimmung der Budgetierung und der freien Mittel dem Management zugesprochen. Dem Arzt bliebe nur die Entscheidung über Annahme oder Ablehnung der Vorgaben des Managements.

Die Realität jedoch sieht meist anders aus. Was den Chefarzt als wichtigen Entscheidungsträger im Spital von einem Abteilungsleiter einer herkömmlichen Firma unterscheidet, ist vor allem seine informellen Macht und sein autonomes Selbstverständnis. Finanzielle Vorgaben seitens des Managements werden in erster Linie als Versuch gewertet, die medizinische Autonomie zu beschneiden. Die Bereitschaft, eine gesamtbetriebliche Sicht einzunehmen und die Vorgaben des Managements als notwendige Massnahmen zur Sicherstellung des gesamten Betriebes zu akzeptieren, ist dagegen oft nur sehr schwach ausgeprägt. Die informelle Macht eines Chefarztes gegenüber dem Management resultiert zum einen aus seiner wissenschaftlichen und medizinischen Kompetenz und dem daraus erwachsenden Ruf, der auch auf das Spital ausstrahlt. Zum anderen verfügen Chefärzte oft über weitreichende Beziehungen in regionalpolitische Kreise und zögern auch nicht, sich dieser Beziehungen zu bedienen, um ihre Interessen durchzusetzen. Nicht zuletzt sind sie häufig Persönlichkeiten, die sich durch Durchsetzungskraft, eine gewisse Aggressivität und ein starkes Selbstvertrauen auszeichnen.

Unter diesem Eindruck scheint es gerechtfertigt, den strikten internen Prinzipal-Agent-Mechanismus mit einem take-it-or-leave-it Angebot des Managements in Frage zu stellen und stattdessen davon auszugehen, dass ein Chefarzt grundsätzlich den Vorschlag des Managements ablehnen und mit einem Gegenvorschlag beantworten wird. Am Universitätsspital Zürich zum Beispiel besteht der Budgetprozess nicht etwa aus einer einmaligen Budgetvorgabe der Finanzabteilung, sondern das Budget wird faktisch im Gegenstromverfahren durch wiederholte Vorschläge und Gegenvorschläge oder gar im Rahmen einer oder mehrerer Verhandlungssitzungen erarbeitet.

Im hier entwickelten Modell wird dieser alternative Entscheidungsmechanismus durch eine asymmetrische *Nash-Verhandlungslösung* umgesetzt, die dem Arzt neben seinem Reservationsnutzen so etwas wie Verhandlungsmacht oder Verhandlungsgeschick gegenüber dem Management zugesteht, um seine Position zu vertreten. Gegenstand der Verhandlungen zwischen Arzt und Management sind die Budgetierung und die freien Mittel, so dass als Ergebnis der Nash-Verhandlungslösung Werte für  $c^d(\theta)$  und  $t^d(\theta)$  resultieren, auf die auch der Arzt gemäss seiner Verhandlungsmacht Einfluss nimmt.

Im Folgenden wird das Modell mit spitalinternen Prinzipal-Agent-Entscheidungsmechanismus

als Modell mit „*PA-Lösung*“ oder „*PA-Mechanismus*“ bezeichnet, die Variante mit der Nash-Verhandlungslösung dagegen als Modell mit „*NV-Lösung*“.

### **5.11 Weitere Annahmen**

In den folgenden Kapiteln werden verschiedene Situationen der Informationsverteilung im Modell betrachtet und analysiert. Weitere Annahmen die für die jeweilige Situation spezifisch sind werden in den jeweiligen Kapiteln getroffen.

## 6 Asymmetrische Information hinsichtlich der Behandlungsqualität

Im Zentrum der Analyse dieses Kapitels steht der Einfluss der beiden spitalinternen Entscheidungsprozesse PA-Lösung und NV-Lösung bei asymmetrischer Information hinsichtlich der Behandlungsqualität bzw. der zur Behandlung eingesetzten Mengen  $q_1$  und  $q_2$  auf das Verhalten des Spitals und auf die daraus resultierende Bedeutung für den Auftraggeber im PA-Modell mit doppeltem Unterstellungsverhältnis.

Zunächst werden jedoch in Abschnitt 6.1.1 die Probleme bei der Messung der Behandlungsqualität in der Praxis erläutert, zudem wird die Annahme der asymmetrischen Information hinsichtlich der Qualität gerechtfertigt. In Abschnitt 6.1.2 folgt als Einstieg in die theoretische Modellierung eine kurze Diskussion zweier Artikel, die sich mit der Problematik der Anreizbildung bei Berücksichtigung der Multidimensionalität der Behandlungsqualität im Rahmen des einfachen PA-Modells beschäftigen.

Unter der Annahme, dass der Auftraggeber die Mengen  $q_1$  und  $q_2$  nicht beobachten kann, wird dann in Abschnitt 6.3 das Modell bei spitalinternem PA-Mechanismus, in Abschnitt 6.4 bei spitalinterner NV-Lösung analysiert. Dabei wird sich herausstellen, dass beide Lösungen zum selben Ergebnis führen, was letztendlich in der Risikoneutralität von Management und Arzt gegenüber der freien Mitteln begründet ist. In den Abschnitten 6.6 und 6.7 erfolgt die Analyse dann unter Annahme eines risikoaversen Arztes.

### 6.1 Praktischer Hintergrund und Abhandlung in der Theorie

#### 6.1.1 Problematische Messbarkeit der Behandlungsqualität in der Praxis

Damit die Behandlungsqualität als Gegenstand der vertraglichen Regelung (zum Beispiel als ein Argument der Vergütung des Spitals) aufgenommen werden kann, müsste sie in irgendeiner Weise beobachtbar und zumindest in irgendeiner Weise messbar sein. Jedoch erweist sich die Messung der Behandlungsqualität in einer Weise, die für den Gebrauch der Messergebnisse als Grundlage für ein Vergütungssystem ausreichend ist, als äusserst problematisch. Dies liegt in erster Linie in der Komplexität und Vielfalt der für die Behandlung erbrachten Leistungen, und dem daraus folgenden *multidimensionalen* Charakter der Behandlungsqualität [vgl. Dranove/Satterwhite (2000), Seite 1133]. Eine vollständige Messung und Bewertung der Behandlungsqualität umfasst im Grunde die Beobachtung und Beurteilung jeder einzelnen im Spital erbrachten Leistung, was natürlich nicht möglich ist.

Daher wird versucht, mit Hilfe von einigen messbaren *Qualitätsindikatoren* Aussagen über



die Qualität der Behandlung zu treffen [vgl. Gowrisankaran (2008)], wobei meist die von Donabedian vorgeschlagene Einteilung der Qualität in *Struktur-, Prozess- und Ergebnisqualität* nachvollzogen wird [Donabedian (1966)]. Die Strukturqualität bezieht sich auf die Qualität der Inputfaktoren und organisatorischen Strukturen, z.B. Management, Informationssysteme oder Qualifikation der Ärzte und des medizinischen Personals. Die Prozessqualität bezieht sich auf die Qualität der Durchführung der Leistungen, z.B. die Einhaltung von Hygiene- und Behandlungsrichtlinien [vgl. McGlynn (2003)]. Die Ergebnisqualität bezieht sich auf das Ergebnis der Behandlung. Entsprechend der Aufteilung in diese drei Bereiche kommen verschiedene Indikatoren zum Einsatz.

Zur Einschätzung der Ergebnisqualität werden häufig Mortalitäts- und Komplikationsraten betrachtet [z.B. Zahnd (2008) oder Bundesamt für Gesundheit (2007)]. Da allerdings der quantitative Zusammenhang zwischen diesen Indikatoren und der Qualität der Leistungserbringung nicht zu ermitteln ist, werden vergleichende Aussagen getroffen, wobei angenommen wird, dass die Qualität dort höher ist, wo beispielsweise die Sterberate geringer ist.

Indikatoren zur Messung der Prozessqualität dagegen setzen direkt bei der Leistungserbringung an. Meistens wird hier die Einhaltung von offiziellen Behandlungsrichtlinien oder die Gabe von Schlüsselmedikationen zur Behandlung eines bestimmten Krankheitsbildes überprüft, beispielsweise die Verabreichung von Acetylsalicylsäure nach einem Herzinfarkt [Fonarow (2009)]. In der Regel werden die entsprechenden Daten von den Spitälern freiwillig geliefert [Lutfiyya (2007)] oder die Informationen werden aus Patientenakten entnommen [Jencks (2000)]. Aus den Ergebnissen der Messung dieser ausgesuchten Indikatoren wird dann auf die Qualität der gesamten Leistungserbringung geschlossen.

Die Strukturqualität ist im Grunde die Voraussetzung für die Prozessqualität [Glickman (2007)]. Soll zum Beispiel die Leistungserbringung der Ärzte den aktuellen medizinischen Erkenntnissen genügen, ist eine ständige Weiterbildung notwendig. Und einen reibungs-freien Ablauf der Behandlung und eine Minimierung von Behandlungsfehlern zu erreichen, sind funktionierende organisatorische Strukturen und Informationssysteme unabdingbar. Zudem hat die Strukturqualität auch einen direkten Einfluss auf den Outcome. Beispiele für Indikatoren der Strukturqualität sind daher Nachweise von Fortbildungen des medizinischen Personals oder der Verfügbarkeit besonderer OP-Methoden wie minimal-invasive Technik [vgl. Qualitätsbericht des Universitätsspitals Zürich (2008)].

Es existieren bei der Messung der Qualität mittels Indikatoren allerdings einige grundsätzliche Probleme, welche die Aussagekraft der Indikatoren und damit ihre Verwendung in der Vergütung der Spitäler in Frage stellen. Zum Beispiel ist eine systematische Bestim-

mung vor allem der Indikatoren, die Struktur- und Prozessqualität abbilden sollen sehr schwierig und teuer, weshalb zur Erhebung dieser Indikatoren oft auf die Selbstauskünfte der Spitäler zurückgegriffen wird. Dies beinhaltet jedoch Anreize zur Falschauskunft.

Ein weiteres Problem besteht darin, dass die zur Verfügung stehenden Struktur- und Prozessindikatoren nur einen Teil des gesamten Spektrums der Leistungserstellung abbilden. Hier besteht die Gefahr, dass sich ein Spital nur auf jene Bereiche konzentriert, die durch Indikatoren abgebildet werden, während Bereiche, für die keine Indikatoren zur Verfügung stehen, vernachlässigt werden.

Bei der Verwendung von Ergebnisindikatoren bestehen ebenfalls einige schwerwiegende Probleme. Zum einen ist der Zusammenhang von Prozessqualität und Ergebnisindikatoren in vielen Bereichen wenig erforscht [Gowrisankaran (2008), Ryan (2009)]. Zum anderen besteht bei den Ergebnisindikatoren das Problem der Risikoadjustierung. Derartige Indikatoren werden auch von Faktoren beeinflusst, die vom Spital nicht zu verantworten sind, so dass eine Adjustierung um diese Faktoren notwendig ist. Zum Beispiel sind Spitäler untereinander bezüglich der behandelten Patienten mitunter sehr heterogen. Grosse Universitätsspitäler behandeln im Durchschnitt Patienten mit schwereren und komplexeren Erkrankungen, bei denen die Komplikationswahrscheinlichkeiten sehr viel höher sind als bei einfacheren Fällen. Bei einem Vergleich der unadjustierten Sterberaten werden die Universitätsspitäler benachteiligt. Erst eine um das erhöhte Risiko korrekt korrigierte Ziffer macht einen Vergleich unter den verschiedenen Spitälern möglich. Jedoch scheinen die zur Verfügung stehenden Datengrundlagen nicht ausreichend, um eine unverzerrte Adjustierung zu ermöglichen. Dementsprechend kritisch sind die Ergebnisse zu betrachten [vgl. Gowrisankaran (2008), Schwappach (2007)]. Ausserdem beinhaltet die Adjustierung für die Spitäler den Anreiz, Patienten nach Eigenschaften zu selektieren, die für die Analysten nicht zu beobachten sind, die jedoch bessere Werte bei den Indikatoren wahrscheinlicher machen [Dranove et al. (2003)].

In Bezug auf die Messung der Qualität und der Verwendung der Ergebnisse zur Bestimmung der Spitalvergütung schliessen wir aus dem geschilderten Sachverhalt:

**Folgerung 6.1** *Zwar kann sich die vom Spital zu verantwortende Qualität in beobachtbaren Indikatoren bemerkbar machen. Jedoch können anhand der verfügbaren Indikatoren keine eindeutigen Rückschlüsse auf die Qualität gemacht werden, so dass die Behandlungsqualität nur unzureichend messbar ist. Dies gilt umso mehr, als dass die Qualität der Behandlung der Vielfältigkeit der zur Behandlung erbrachten Leistungen und der notwendigen Koordination entsprechend mehrdimensional ist.*

### 6.1.2 Asymmetrische Information hinsichtlich der Qualität in der theoretischen Literatur zur Spitalregulierung

Die aus der schwierigen Messbarkeit der Qualität folgende Problematik der allokativ ineffizienten Qualitätsbereitstellung durch die Spitäler wird in einer Reihe von Beiträgen thematisiert, in denen die Analyse im Rahmen des einfachen PA-Modells durchgeführt wird [z.B. Jack (2005), Eggleston (2005), Chalkley/Malcomson (1998a + 1998b)]. Eggleston (2005) zum Beispiel greift in ihrem Beitrag das Problem der unvollständigen Erfassung aller Dimensionen der Qualität durch Indikatoren auf. In ihrem Modell versucht der Auftraggeber mittels eines Pay-for-Performance-Anreizsystems, der finanziellen Berücksichtigung von beobachtbaren Qualitätsindikatoren, das Spital in der Bereitstellung der Qualität zu beeinflussen.<sup>40</sup> Da aber die ihm zur Verfügung stehenden Indikatoren nur einen Teil der Qualitätsbereitstellung abdecken, besteht die Gefahr, dass das Spital sich auf jene Dimensionen der Qualität konzentriert, die mittels Indikatoren gemessen und bei der Vergütung berücksichtigt werden. Eggleston zeigt am Beispiel eines zweidimensionalen Qualitätsvektors, von dem nur eine Dimension über Indikatoren darstellbar ist, dass der Auftraggeber mittels Pay-for-Performance die Bereitstellung dieser Dimension durch das Spital verbessern kann. Allerdings geht dies mit einer niedrigeren Qualität in der anderen, nicht beobachtbaren Dimension einher, da Eggleston annimmt, dass der Nutzenverlust, der dem Spital aus der Bereitstellung entsteht, in beiden Dimensionen ansteigt und konvex ist. Dieser Effekt ist umso ausgeprägter, je mehr Kostenverantwortung das Spital tragen muss, so dass in dem Modell ein gemischtes System mit nur teilweiser Kostenverantwortung für den Auftraggeber das beste Ergebnis ermöglicht. Die first best-Allokation ist im Allgemeinen jedoch nicht möglich.<sup>41</sup>

Im Modell von Chalkley/Malcomson (1998b) verfügt der Auftraggeber nicht über die

---

<sup>40</sup>Solche Pay-for-Performance-Systeme werden vor allem in den USA eingesetzt. Zum Beispiel basiert das 2003 gestartete Programm CMS-Premier Hospital Quality Demonstration, ein gemeinsames Projekt der US-Bundesbehörde Centers for Medicare and Medicaid Services und der Non-Profit-Organisation Premier, einem Netzwerk von Spitälern in den USA, auf 33 Prozess- und Outcomeindikatoren, die im Zusammenhang mit den fünf Krankheitsbildern Herzversagen, Herzinfarkt, Lungenentzündung, Koronararterien-Bypass und künstliche Hüft- und Kniegelenke stehen. Aufgrund der Indikatoren werden Ranglisten der teilnehmenden Spitäler erstellt und je nach Platzierung erhält ein Spital einen Bonus oder einen Malus [Ryan (2009)]. Allerdings ist die Effektivität des Programms sowie von Pay-for-Performance als solches empirisch umstritten. Zwar wurden leichte Verbesserungen bei Indikatoren festgestellt [Mehrotra (2009)], jedoch nicht bei der Outcomequalität [Glickman (2007)]. Zudem besteht die oben genannte Gefahr, dass sich das Spital verstärkt auf die Bereiche konzentriert, die von den Indikatoren dargestellt werden, während übrige Bereiche eventuell vernachlässigt werden. Und da die Prozessindikatoren vom Spital gemeldet werden, sind verzerrte Darstellungen nicht auszuschließen.

<sup>41</sup>In diesem Sinne greift hier Tinbergens's Gesetz, nachdem der Auftraggeber über genauso viele Instrumente verfügen muss wie er Ziele verfolgt, um bei allen Zielen die optimale Ausprägung zu erreichen [vgl. Rosado (2004)].

Möglichkeit, mittels Pay-for-Performance die Qualitätsbereitstellung zu beeinflussen. Allerdings nehmen Chalkley und Malcomson an, dass sich die Bereitstellung der Qualität in allen ihren Dimensionen in der Anzahl der Patienten widerspiegelt, die Behandlung im Spital nachfragen. Wenn die Nachfrage für den Auftraggeber zu beobachten ist, dann kann der Auftraggeber diese Information gleichsam als Pay-for-Performance-Indikator verwenden, um Einfluss auf die Qualität zu nehmen. Allerdings hat der Auftraggeber auch hier das Problem, dass er mit den beobachtbaren Kosten und der Nachfrage nicht alle Dimensionen der Qualitätsbereitstellung optimal steuern kann. Ein Sonderfall ergibt sich jedoch, wenn die Patienten die Relationen der verschiedenen Dimensionen der Qualität zueinander gleich bewerten wie der Auftraggeber. Dann muss der Auftraggeber lediglich dafür sorgen, dass das Spital die Mittel für die Bereitstellung der Qualität insgesamt richtig wählt; der optimale Mix wird durch die Nachfrage sichergestellt. In diesem Fall erreicht ein System, welches die volle Kostenverantwortung dem Spital überträgt, die first best-Allokation. Die Anreize zur optimalen Bereitstellung der Mittel für die Qualität gibt der Auftraggeber dem Spital mit einer Pauschale für das Spital für jeden Patient, der die Leistungen nachfragt.<sup>42</sup>

## **6.2 Asymmetrische Information hinsichtlich der Behandlungsqualität im Modell mit doppeltem Unterstellungsverhältnis**

### **6.2.1 Annahmen bezüglich der Informationsverteilung**

Wir nehmen an, dass der Auftraggeber weder die Menge  $q_1$  noch  $q_2$  beobachten kann. Auch nehmen wir an, dass eventuell erhobene Indikatoren qualitativ nicht die notwendigen Anforderungen erfüllen, um als Argument der Spitalvergütung Verwendung finden können.

Während die Probleme des Auftraggebers bei der Beobachtung der Qualität bereits dargestellt wurden, ist die asymmetrische Information bezüglich der Qualität zwischen Arzt und Management dagegen nicht unbedingt eindeutig gegeben. Auf der einen Seite ist die Behandlung in einem Spital eine sehr komplexe Angelegenheit, was eine Informationsverarbeitung für das Management sehr schwer macht, vor allem wenn man bedenkt, dass in einem grossen Spital 30'000 und mehr stationäre Fälle pro Jahr behandelt werden. Auf der anderen Seite finden mehr und mehr EDV-gestützte Analysesysteme Verwendung im Spitalmanagement, die auch die Generierung und Verarbeitung von grossen Datenmengen möglich machen [Zietz (2011)], während die kontinuierliche Entwicklung der Kos-

---

<sup>42</sup>Tatsächlich ist das Modell von Chalkley und Malcomson etwas komplexer, da die Autoren zusätzlich die Kosteneffizienz als weiteres Problem betrachten. Ausserdem lassen sie die Möglichkeit zu, dass die Anzahl der Patienten, die Behandlung nachfragen, nicht gleich der tatsächlich behandelten entspricht.

tenrechnungssysteme ein relativ differenziertes Bild des internen Ressourcenverbrauchs ermöglicht. Der Aufwand für die verschiedenen medizinischen Materialien, Medikamente, Personalkosten usw. werden beispielsweise in Schweizer Spitälern gemäss offiziellen Rechnungslegungsvorschriften auf separate Kostenarten gebucht, so dass das Management zumindest die Kosten der einzelnen Massnahmen voneinander trennen kann. In unserem Modell wäre es also eine plausible Annahme, dass das Management  $p_1q_1$  und  $p_2q_2$  beobachten kann. In Kenntnis der Preise könnte das Management dann die Werte von  $q_1$  und  $q_2$  herleiten.

Doch auch wenn das Management die einzelnen Mengen beobachten kann, ist fraglich, ob das Management diese Information tatsächlich nutzen könnte, um dem Arzt die Mengen vorzuschreiben. Zwar kann das Management die Rahmenbedingungen für das medizinische Handeln vorgeben, hat aber nur beschränkte Einflussmöglichkeiten auf die medizinischen Prozesse und die operative Leistungserbringung innerhalb der Fachabteilungen [Bunzemeier (2010)], zumal das Management dem Arzt in Behandlungsfragen nicht weisungsbefugt ist. Ausserdem wird das Management nicht die notwendige Fachkompetenz haben, um dem Arzt als Experten auf seinem Fachgebiet konkrete Therapievorschriften machen zu können und zu wollen, zumal es - zumindest in diesem Modell - sowieso keine Präferenzen hinsichtlich der Zusammensetzung der Behandlungsmengen hat.

In unserem Modell gehen wir deshalb davon aus, dass das Management keine Anstrengungen unternimmt, die Mengen  $q_1$  und  $q_2$  zu beobachten. Das primäre Interesse des Managements bei der Behandlung eines Patienten durch den Arzt ist die Einhaltung der Kostenbudgetierung.

Die Preise  $p_1$  und  $p_2$  sowie die Kosten der Behandlung  $c$  sind dagegen von allen Beteiligten zu beobachten.

In Bezug auf die Fallschwere  $\theta$  nehmen wir in diesem Kapitel an, dass der Arzt, nachdem er die Fallschwere des Patienten festgestellt hat, diese dem Management berichtet. Der Bericht des Arztes an das Management ist vom Auftraggeber nicht zu beobachten, so dass der Auftraggeber auf die Weiterleitung des Berichtes durch das Management angewiesen ist. In diesem Kapitel gehen wir davon aus, dass der Arzt als auch das Management *wahrheitsgemäss* berichten. Die Problematik der versteckten Information, die sich ergibt, wenn diese Annahme fallengelassen wird, betrachten wir in Kapitel 8.

Tabelle 1 gibt einen Überblick über die Verteilung der Information in Kapitel 6.

<i>Grösse beobachtbar ja/nein</i>	Auftraggeber	Management	Arzt
Mengen $q_1, q_2$	nein	nein	ja
Preise $p_1, p_2$	ja	ja	ja
Kosten $c$	ja	ja	ja
Fallschwere $\theta$	ja	ja	ja

Tabelle 1: Informationsverteilung des Modells in Kapitel 6.

### 6.2.2 First best-Allokation

Um den Effekt der asymmetrischen Informationsverteilung hinsichtlich der Quantitäten auf die bestmögliche Allokation aus Sicht des Auftraggebers identifizieren zu können, wird zunächst die first best-Allokation hergeleitet. Die Allokation also, die der Auftraggeber wählen würde, wenn er sämtliche relevanten Variablen beobachten und vertraglich regeln bzw. einklagen könnte. Die first best-Allokation gilt als Referenz für die PA-Lösung als auch für die NV-Lösung.

Im Falle von vollständiger Information gibt der Auftraggeber die aus seiner Sicht optimalen Quantitäten  $q_1^{FB}(\theta)$  und  $q_2^{FB}(\theta)$  für die Behandlung eines Patienten der Fallschwere  $\theta$  vor und vergütet das Management mit der Pauschale  $k(\theta)$ . Sein Maximierungsprogramm lautet

$$\max_{q_1(\theta), q_2(\theta), k(\theta)} W(\theta) = \theta B(.) - k(\theta) - (p_1 q_1(\theta) + p_2 q_2(\theta)), \quad (40)$$

wobei  $\theta B(.) \equiv \theta B(q_1(\theta), q_2(\theta))$ . Der Ausdruck (40) lässt sich äquivalent auch in Abhängigkeit des gemeinsamen Überschusses von Management und Arzt

$$N\ddot{U}(\theta) = G(\theta) + U(\theta) - u = k(\theta) + \theta V(.) - u, \quad (41)$$

mit  $\theta V(.) \equiv \theta V(q_1(\theta), q_2(\theta))$ , formulieren:

$$\max_{q_1(\theta), q_2(\theta), N\ddot{U}(\theta)} W(\theta) = \theta B(.) + \theta V(.) - (p_1 q_1(\theta) + p_2 q_2(\theta)) - N\ddot{U}(\theta) - u. \quad (42)$$

Für die Bedingungen erster Ordnung für die optimalen Mengen  $q_1^{FB}(\theta)$  und  $q_2^{FB}(\theta)$  erhalten wir

$$\theta B_{q_1}^{FB}(.) + \theta V_{q_1}^{FB}(.) = p_1 \quad (43)$$

$$\theta B_{q_2}^{FB}(.) + \theta V_{q_2}^{FB}(.) = p_2, \quad (44)$$

wobei  $\theta B_{q_1}^{FB}(\cdot) \equiv \theta B_{q_1}^{FB}(q_1^{FB}(\theta), q_2^{FB}(\theta))$  usw.

Der gemeinsame Nutzenüberschuss  $N\ddot{U}(\theta)$  geht eindeutig negativ in die Zielfunktion (42) ein, so dass der Auftraggeber für den Überschuss den minimalen Wert  $g$  wählen wird, so dass die externe Teilnahmebedingung des Managements (38) gerade bindet. Bei der first best-Allokation erhalten Arzt und Management also mit

$$N\ddot{U}(\theta) = g \tag{45}$$

den minimalen Nutzenüberschuss. Wie der Nutzenüberschuss allerdings zwischen Arzt und Management aufgeteilt wird, spielt für die first best-Allokation keine Rolle, solange die Teilnahme von Arzt und Management gewährleistet ist.

### 6.3 Analyse des Modells bei asymmetrischer Information und spitalinterner Prinzipal-Agent-Lösung

Im Folgenden wird das Modell unter der Annahme analysiert, dass zwischen Management und Arzt ein PA-Verhältnis herrscht und das Management die Fallbudgetierung  $c^d(\theta)$  und die freien Mittel  $t^d(\theta)$  autonom bestimmt. Zunächst legen wir den zeitlichen Ablauf des Modells in dieser Situation fest.

#### 6.3.1 Zeitliche Abfolge

Der zeitliche Ablauf des Spiels ist wie folgt:

1. In der ersten Stufe des Spieles bestimmt der Auftraggeber die Parameter  $k(\theta)$  und  $r(\theta)$  für jede Fallschwere  $\theta \in \Theta$ .
2. In der zweiten Stufe entscheidet das Management über die Annahme oder die Ablehnung des Vertragsangebotes des Auftraggebers. Wenn das Management annimmt, bestimmt es für jede Fallschwere  $\theta \in \Theta$  die Fallbudgetierung  $c^d(\theta)$  und die freien Mittel  $t^d(\theta)$ .
3. In der dritten Stufe kommen die  $n$  Patienten zur Behandlung in das Spital. Der Arzt stellt die jeweilige Fallschwere  $\theta$  der einzelnen Patienten fest und entscheidet, ob er den jeweiligen Patienten zur Behandlung aufnimmt oder ob er ihm die Aufnahme und die Behandlung verweigert. Wenn sich der Arzt für die Behandlung des Patienten entscheidet, berichtet er die Fallschwere wahrheitsgemäss an das Management. Das Management entscheidet, ob es die Behandlung zulässt. Wenn das Management

die Behandlung zulässt, wählt der Arzt die Mengen  $q_1, q_2$  für die Behandlung des Patienten so, dass das vom Management vorgegebene Budget  $c^d(\theta)$  eingehalten wird. Das Management berichtet die Fallschwere wahrheitsgemäss an den Auftraggeber. Die Behandlungskosten werden von allen drei Akteuren beobachtet. Die Zahlungen finden statt.

Wie bei Spielen mit endlicher Anzahl von Stufen üblich, wird die Analyse mittels Rückwärts-Induktion angegangen. Zuerst wird also das Verhalten des Arztes bei gegebenen Entscheidungen von Management und Auftraggeber untersucht.

### 6.3.2 Bestimmung der Mengen $q_1$ und $q_2$ durch den Arzt

Der Arzt wählt bei der Behandlung eines Patienten der Fallschwere  $\theta$  die Quantitäten  $q_1(\theta)$  und  $q_2(\theta)$ , muss dabei allerdings das vom Management definierte Budget  $c^d(\theta)$  einhalten. Das Maximierungsproblem des Arztes bei der Behandlung eines Patienten der Fallschwere  $\theta$  lautet also

$$\max_{q_1(\theta), q_2(\theta)} U(\theta) = t^d(\theta) + \theta V(q_1(\theta), q_2(\theta)) \quad (46)$$

unter der Bedingung

$$c^d(\theta) = p_1 q_1(\theta) + p_2 q_2(\theta). \quad (47)$$

Wenn die Budgetbedingung nicht bindet, ist der Arzt in seiner Wahl nicht eingeschränkt, und er behandelt den Patienten mit den Sättigungsmengen  $q_1^{od}, q_2^{od}$ , also jenen Mengen, bei denen sein Grenznutzen aus der Behandlung gleich null ist.

Bindet die Budgetrestriktion, und davon gehen wir hier grundsätzlich aus, dann entspricht das Maximierungsproblem jenem des Haushaltes bei der Wahl eines Güterbündels unter Budgetbeschränkung aus der Konsumententheorie, so dass wir hier die entsprechenden Ergebnisse verwenden können [z.B. Breyer (2004), Seite 131].

Die Ableitungen der zum Programm des Arztes zugehörigen Lagrange-Funktion liefern die Bedingungen für ein Maximum:

$$\theta V_{q_1}(q_1^*(\theta), q_2^*(\theta)) = \lambda^* p_1 \quad (48)$$

$$\theta V_{q_2}(q_1^*(\theta), q_2^*(\theta)) = \lambda^* p_2 \quad (49)$$

$$c^d(\theta) = p_1 q_1^*(\theta) + p_2 q_2^*(\theta), \quad (50)$$



wobei  $\lambda$  der Lagrange-Multiplikator des Maximierungsproblems ist. Der hochgestellte Index  $*$  steht dabei für die optimalen Werte. Zur übersichtlicheren Darstellung verwenden wir im Folgenden die Notation  $\theta V_{q_h}^*(.) \equiv \theta V_{q_h}(q_1^*(\theta), q_2^*(\theta))$ ,  $h = 1, 2$  usw.

Da  $\theta V(.)$  per Annahmen streng konkav und die Beschränkung (47) linear in den Mengen  $q_1$  und  $q_2$  ist, identifiziert das Gleichungssystem (48) - (50) ein globales Maximum [Hoy et al. (2001), Theorem 13.8]. Die angenommene Linearität der Bewertungsfunktion in der Fallschwere dagegen führt hier dazu, dass die optimalen Mengen nur indirekt über die Budgetierung  $c^d(\theta)$  von der Fallschwere abhängen (s. Anhang A.1).<sup>43</sup> Es gilt also:

$$q_1^*(\theta) \equiv q_1^*(c^d(\theta)) \quad (51)$$

$$q_2^*(\theta) \equiv q_2^*(c^d(\theta)). \quad (52)$$

Der Wert des Lagrange-Multiplikators im Optimum hängt dagegen auch direkt von der Fallschwere ab:

$$\lambda^* \equiv \lambda^*(\theta, c^d(\theta)). \quad (53)$$

Um die Darstellung übersichtlich zu halten, verwenden wir im Folgenden bisweilen die vereinfachende Notation  $q_1^*(.) = q_1^*(c^d(\theta))$ ,  $q_2^*(.) = q_2^*(c^d(\theta))$  und  $\lambda^*(\theta, .) = \lambda^*(\theta, c^d(\theta))$ .

**6.3.2.1 Anpassung der Mengen an eine Budgetänderung** Entscheidend für das Management bei der Bestimmung der Budgetvorgabe ist natürlich, wie der Arzt auf eine Änderung der Vorgabe reagiert. Aus der Konsumententheorie wissen wir, dass eine Lockerung der Budgetrestriktion nicht unbedingt mit einer Zunahme beider Mengen  $q_1^*(c^d(\theta))$  und  $q_2^*(c^d(\theta))$  verbunden sein muss. So reagiert der Arzt auf eine Budgetausweitung mit einer Reduktion der entsprechenden Menge, wenn eine Behandlungsleistung für den Arzt die Bedeutung eines inferioren Gutes hat, was durchaus denkbar ist. Mit der Annahme, dass sich beide Leistungen in ihrer Wirkung ergänzen (Ausdruck (34)), schliessen wir diese Möglichkeit jedoch aus.

Da die Budgetrestriktion für den Arzt bindet, erhalten wir mit totalem Differenzieren der Bedingungen (48) - (50) und unter Anwendung der Cramer'schen Regel [Hoy et al. (2001), S. 654] für die Reaktionen des Arztes auf eine marginale Anpassung der Budgetierung  $c^d(\theta)$ :

---

<sup>43</sup>Dies bedeutet, dass der Arzt bei gegebener Budgetierung für alle Fallschweren den gleichen Mix der beiden Massnahmen einsetzt. Wie wir weiter unten sehen werden, ist jedoch die Zahlungsbereitschaft des Arztes für eine Ausweitung des Budgets und damit für eine Ausweitung der Mengen umso grösser, je höher die Fallschwere ist.

$$\frac{\partial q_1^*(c^d(\theta))}{\partial c^d(\theta)} = -\frac{1}{|H|}[p_1\theta V_{q_2q_2}^*(.) - p_2\theta V_{q_2q_1}^*(.)] \quad (54)$$

$$\frac{\partial q_2^*(c^d(\theta))}{\partial c^d(\theta)} = -\frac{1}{|H|}[p_2\theta V_{q_1q_1}^*(.) - p_1\theta V_{q_2q_1}^*(.)], \quad (55)$$

wobei  $|H| = 2\theta V_{q_1q_2}^* p_1 p_2 - \theta V_{q_1q_1}^* p_2^2 - \theta V_{q_2q_2}^* p_1^2 > 0$  die Determinante der Hesse Matrix aus dem Maximierungsproblem des Arztes ist. Da das Maximierungsproblem des Arztes ein eindeutiges Maximum hat, ist die Determinante überall positiv. Mit der Annahme  $V_{q_1q_2}(\cdot) \geq 0$  folgt

$$\frac{\partial q_h^*(.)}{\partial c^d(\theta)} > 0, \quad h = 1, 2. \quad (56)$$

Der Arzt reagiert auf eine Budgetausweitung mit einer Ausweitung der beiden Mengen  $q_1^*(c^d(\theta))$  und  $q_2^*(c^d(\theta))$ .<sup>44</sup>

### 6.3.3 Zahlungsbereitschaft des Arztes für eine Ausweitung des Budgets

Geht man davon aus, dass die Budgetbedingung (47) bindet und den Arzt bei der Wahl der Behandlungsmengen einschränkt, dann bedeutet bei gegebenen freien Mitteln  $t^d(\theta)$  eine marginale Budgetausweitung für den Arzt eine Erhöhung des maximal erreichbaren Nutzens. Dieser marginale Zuwachs gemessen in Geldeinheiten entspricht seiner Zahlungsbereitschaft für eine marginale Budgeterhöhung.

Die Zahlungsbereitschaft des Arztes für eine marginale Lockerung der Budgetierung lässt sich leicht bestimmen. Mit dem Envelope-Theorem entspricht sie dem Wert des Lagrange-Multiplikators  $\lambda^*(\theta, c^d(\theta))$  [vgl. z.B. Mas-Colell/Whinston/Green (1995), Seite 54]. Für den Wert des Lagrange-Multiplikators im Optimum  $\lambda^*(\theta, c^d(\theta))$  folgt aus den notwendigen Bedingungen (48) - (50)

$$\lambda^*(\theta, c^d(\theta)) = \frac{\theta V_{q_1}^*(.)}{p_1} = \frac{\theta V_{q_2}^*(.)}{p_2} > 0. \quad (57)$$

Die Zahlungsbereitschaft des Arztes für eine Budgetausweitung ist positiv, da  $\theta V_{q_1}^*(.)$  bzw.  $\theta V_{q_2}^*(.)$  grösser null ist.

Die Zahlungsbereitschaft ist im Allgemeinen nicht konstant und ändert sich mit der Höhe der Budgetierung und der Fallschwere. Ableiten von  $\lambda^*(\theta, c^d(\theta))$  nach  $c^d(\theta)$  ergibt mit (57)

---

<sup>44</sup>Im Sinne der Konsumententheorie haben die beiden Leistungen für den Arzt die Bedeutung von normalen Gütern und die Engel-Kurven beider Massnahmen weisen eine strikt positive Steigung auf.

$$\frac{\partial \lambda^*(\theta, \cdot)}{\partial c^d(\theta)} = \frac{\theta}{p_1} \left[ V_{q_1 q_1}^*(\cdot) \frac{\partial q_1^*(\cdot)}{\partial c^d(\theta)} + V_{q_1 q_2}^*(\cdot) \frac{\partial q_2^*(\cdot)}{\partial c^d(\theta)} \right] \quad (58)$$

Mit (54) und (55) ergibt sich

$$\frac{\partial \lambda^*(\theta)}{\partial c^d(\theta)} \leq 0 \leftrightarrow V_{q_1 q_2}^*(\cdot)^2 \leq V_{q_1 q_1}^*(\cdot) V_{q_2 q_2}^*(\cdot) \quad (59)$$

was wegen der Konkavität von  $V(\cdot)$  der Fall ist (Ausdruck (33)).

Eine marginale Erhöhung der Budgetierung hat auf die ärztliche Bewertung der marginalen Mengen grundsätzlich zwei gegensätzliche Effekte. Auf der einen Seite sinkt durch die marginale Ausweitung der Mengen wegen der abnehmenden Grenzwirksamkeit der Massnahmen der marginale Nutzen aus den beiden Mengen. Auf der anderen Seite führt die angenommene Ergänzung der beiden Massnahmen ( $V_{q_1 q_2}(\cdot) > 0$ ) zu einer Erhöhung des marginalen Nutzens beider Massnahmen. Je nach Ausprägung dominiert der eine oder andere Effekt, so dass beides möglich ist: Die Zahlungsbereitschaft sinkt oder steigt mit dem Budget. Ergänzen sich zum Beispiel beide Massnahmen in einem extremen Ausmass und steigert der Einsatz einer Leistung die Wirksamkeit der anderen so sehr, dass die an sich abnehmende Produktivität der beiden Leistungen allein mehr als ausgeglichen wird, dann steigt die Zahlungsbereitschaft des Arztes für eine Budgetausweitung mit dem Budget.

Durch die Annahme der strikten Konkavität der Bewertungsfunktion  $V(\cdot)$  dominiert der Effekt der abnehmenden Wirksamkeit, so dass die Zahlungsbereitschaft für eine Budgetausweitung mit dem Budget abnimmt. Da sie jedoch bei null anlangt, wenn das Budget ausreicht, um die Sättigungsmenge des Arztes zu finanzieren, ist die zusätzliche Annahme einer konkaven Funktion gerechtfertigt:

$$\frac{\partial \lambda^*(\cdot)}{\partial c^d(\theta)} \leq 0, \quad \frac{\partial^2 \lambda^*(\theta, c^d(\theta))}{(\partial c^d(\theta))^2} < 0. \quad (60)$$

Dagegen nimmt die Zahlungsbereitschaft (bei konstanter Budgetierung) linear mit der Fallschwere zu. Ableiten von  $\lambda^*(\theta, c^d(\theta))$  nach  $\theta$  ergibt mit (57)

$$\frac{\partial \lambda^*(\cdot)}{\partial \theta} = \frac{V_{q_1}^*(\cdot)}{p_1} > 0. \quad (61)$$

Die Linearität folgt, da die Mengen  $q_1^*(c^d(\theta))$  und  $q_2^*(c^d(\theta))$  nicht direkt von der Fallschwere abhängen.

#### 6.3.4 Bestimmung der Budgetierung $c^d(\theta)$ und der freien Mittel $t^d(\theta)$ durch das Management.

In der zweiten Stufe des Spieles bestimmt das Management die freien Mittel  $t^d(\theta)$  und die Budgetierung  $c^d(\theta)$  für jede Fallschwere  $\theta \in \Theta$  bei gegebener, vom Auftraggeber definierter, Pauschale  $k(\theta)$  und Kostenbeteiligung  $r(\theta)$ . Dabei muss es die Anreizverträglichkeitsbedingungen und die Teilnahmebedingung (36) des Arztes beachten. Die Anreizverträglichkeitsbedingungen werden durch die aus der Budgetierung  $c^d(\theta)$  resultierende Reaktion des Arztes  $q_1^*(c^d(\theta))$  und  $q_2^*(c^d(\theta))$  dargestellt. Für das Maximierungsproblem des Managements folgt damit pro Patient der Fallschwere  $\theta$

$$\max_{t^d(\theta), c^d(\theta)} G(\theta) = k(\theta) - r(\theta)c^d(\theta) - t^d(\theta) \quad (62)$$

s.t.

$$t^d(\theta) \geq u - \theta V^*(.), \quad (63)$$

mit  $\theta V^*(.) = \theta V(q_1^*(c^d(\theta)), q_2^*(c^d(\theta)))$  dem aus der Budgetierung resultierenden Behandlungsnutzen für den Arzt. Da die freien Mittel eindeutig negativ in die Zielfunktion (62) des Managements eingehen, wird es die freien Mittel so niedrig wählen, dass die Teilnahmebedingung (36) des Arztes bindet. Damit sind die freien Mittel mit der Bestimmung der Budgetierung  $c^d(\theta)$  gegeben. Einsetzen der bindenden Teilnahmebedingung (63) in (62) führt zu dem äquivalenten Maximierungsproblem

$$\max_{c^d(\theta)} G(\theta) = k(\theta) - r(\theta)c^d(\theta) + \theta V^*(.) - u. \quad (64)$$

Diese Gleichung stellt die aus Sicht des Auftraggebers effektive Zielfunktion des Spitals dar.

Aus dem vorigen Abschnitt wissen wir, dass die marginale Erhöhung des Behandlungsnutzens  $\theta V^*(.)$  infolge einer marginalen Lockerung der Budgetierung dem Lagrange-Multiplikator  $\lambda^*(.)$  aus dem Maximierungsproblem des Arztes entspricht:

$$\frac{\partial \theta V^*(.)}{\partial c^d(\theta)} = \lambda^*(\theta, c^d(\theta)). \quad (65)$$

Damit ergibt sich für die notwendige Bedingung für das optimale Budget  $c^{d*}(\theta)$ :

$$\lambda^*(\theta, c^d(\theta)) - r(\theta) = 0. \quad (66)$$

Gleichung (66) hat folgende Interpretation: Der marginale Nutzenzuwachs für den Arzt aus einer marginalen Erhöhung der Budgetierung beträgt  $\lambda^*(\theta, c^d(\theta))$ . Der Arzt ist also indifferent gegenüber einer marginalen Erhöhung des Budgets, wenn das Management gleichzeitig die freien Mittel um den Betrag  $\lambda^*(\theta, c^d(\theta))$  senkt. Für das Management ist es optimal, wenn der Arzt immer gerade seinen Reservationsnutzen  $u$  erhält. Um den Nutzen des Arztes bei einer marginalen Erhöhung der Budgetierung auf dem Niveau des Reservationsnutzens zu halten, kann das Management die freien Mittel  $t^d(\theta)$  um  $-\lambda^*(\theta, c^d(\theta))$  senken. Auf der anderen Seite muss das Management bei einer marginalen Kostenausweitung den Anteil  $r(\theta)$  selber tragen. Im Optimum wählt es die Budgetierung derart, dass der Anteil  $r(\theta)$  an den marginalen Kosten der durch die marginale Budgetausweitung möglichen Reduktion der freien Mittel  $t^d(\theta)$  entspricht.

Wegen (60) ist die Bedingung zweiter Ordnung

$$\frac{\partial^2 \theta V^*(.)}{(\partial c^d(\theta))^2} = \frac{\partial \lambda^*(\theta, c^d(\theta))}{\partial c^d(\theta)} < 0 \quad (67)$$

erfüllt.

Aus (66) lässt sich das optimale Budget  $c^{d*}(\theta)$  für einen Patienten der Fallschwere  $\theta$  als eine Funktion der vom Auftraggeber definierten Kostenbeteiligung schreiben:

$$c^{d*}(\theta) \equiv c^{d*}(\theta, r(\theta)). \quad (68)$$

Da es für das Management optimal ist, wenn die Teilnahmebedingung des Arztes bindet, sind die optimalen freien Mittel  $t^{d*}(\theta)$  mit der Wahl der Budgetierung gegeben und daher ebenfalls eine Funktion der Kostenbeteiligung  $r(\theta)$ :

$$t^{d*}(\theta, r(\theta)) = u - \theta V^*(c^{d*}(\theta, r(\theta))), \quad (69)$$

wobei  $\theta V^*(c^{d*}(\theta, r(\theta)))$  der Behandlungsnutzen des Arztes bei der Budgetierung  $c^{d*}(\theta, r(\theta))$  ist.

### 6.3.5 Anreizwirkung der Kostenbeteiligung $r(\theta)$

Mit dem Kostenbeteiligungsparameter  $r(\theta)$  kann der Auftraggeber durch Übertragung von Kostenverantwortung über das Management Einfluss auf die Allokation der Mengen nehmen, ohne dem Spital eine konkrete Budgetierung vorschreiben zu müssen.

Es leuchtet intuitiv ein, dass das Management eher bereit ist, dem Arzt ein engeres Budget vorzugeben, wenn seine Kostenverantwortung höher ist, selbst wenn das Management dafür dem Arzt einen relativ hohen Betrag an freien Mitteln zur Verfügung stellen muss.

Bei einer Kostenbeteiligung von 100% ( $r(\theta) = 1$ ) wählt das Management ein derart niedriges Budget  $c^{d*}(\theta)$ , dass die Zahlungsbereitschaft des Arztes den Wert  $\lambda^*(\theta, c^{d*}(\theta)) = 1$  annimmt, d.h. die Kosten der Behandlung werden vollständig vom Spital internalisiert, wobei entsprechend geringe Mengen für die Behandlung resultieren.

Ist die Kostenbeteiligung dagegen niedrig, haben die Kosten gegenüber den freien Mitteln eine geringere Bedeutung, so dass das Management bei der Budgetierung eher grosszügig ist, um die freien Mittel niedrig zu halten. Hat das Management überhaupt keine Kostenverantwortung zu tragen ( $r(\theta) = 0$ ), wählt es eine Budgetierung die mindestens so hoch ist, dass die Zahlungsbereitschaft des Arztes für eine Budgetausweitung bei  $\lambda^*(\theta, \cdot) = 0$  ist, d.h. dass der Arzt den Patienten mit den Sättigungsmengen behandeln kann.

Formal ergibt sich durch implizite Differenzierung der Bedingung erster Ordnung (66) für die Reaktion des Managements auf eine Erhöhung der Kostenbeteiligung

$$\frac{dc^{d*}(\theta, r(\theta))}{dr(\theta)} = \frac{1}{\partial \lambda^*(\theta, \cdot) / \partial c^d(\theta, r(\theta))} < 0. \quad (70)$$

Mit der aus einer höheren Kostenverantwortung des Managements resultierenden Budgetreduktion beim Arzt geht wegen (56) eine Reduktion beider Mengen einher.

Für den Einfluss auf die freien Mittel ergibt sich

$$\frac{dt^{d*}(\theta, r(\theta))}{dr(\theta)} = \lambda^*(\theta, c^{d*}(\theta)) > 0, \quad (71)$$

da der Arzt für die Budgetkürzung mit einer Erhöhung der freien Mittel um  $\lambda^*(\theta, c^{d*}(\theta))$  Geldeinheiten kompensiert werden muss.

Eine Änderung der Pauschale  $k(\theta)$  hat dagegen keine Auswirkung auf die Kosten, die Mengen oder die freien Mittel.

### 6.3.6 Optimale Pauschale $k(\theta)$ und Kostenbeteiligung $r(\theta)$ aus Sicht des Auftraggebers

Der Auftraggeber steht jetzt vor dem Problem, mit der Bestimmung der Parameter  $k(\theta)$  und  $r(\theta)$  für jede Fallschwere die jeweiligen optimalen Mengen  $q_1(\theta)$  und  $q_2(\theta)$  zu errei-

chen. Da die Pauschale  $k(\theta)$  keine Anreizwirkung entfaltet und keinen Einfluss auf die spitalinterne Budgetierung und damit auf die im Spital bestimmten Mengen  $q_1^*(.)$  und  $q_2^*(.)$  hat, ist der Auftraggeber insbesondere mit dem Problem konfrontiert, mit dem einen Parameter  $r(\theta)$  die Bestimmung der beiden Mengen zu beeinflussen. Mit der Festlegung der Beteiligung  $r(\theta)$  kann der Auftraggeber dem Management lediglich den Anreiz geben, dem Arzt bei der Behandlung eines Patienten der Fallschwere  $\theta$  ein bestimmtes Budget vorzugeben. Die Zusammensetzung der Mengen zur Behandlung bestimmt der Arzt aber gemäss seinen eigenen Präferenzen unter Einhaltung der Budgetvorgabe. Wenn sich die Präferenzen des Arztes von denen des Auftraggebers (in der marginalen Bewertung der Quantitäten) unterscheiden, dann sind aus Sicht des Auftraggebers Verzerrungen nicht zu vermeiden. Die Pauschale  $k(\theta)$  hat hier lediglich die Funktion, die Höhe des Nutzenüberschusses  $N\ddot{U}(\theta)$  von Management und Arzt zu steuern.

Das formelle Maximierungsprogramm des Auftraggebers bei einem Patienten der Fallschwere  $\theta$  lautet<sup>45</sup>

$$\max_{k(\theta), r(\theta)} W = \theta B^*(.) - (1 - r(\theta))c^{d*}(\theta, r(\theta)) - k(\theta) \quad (72)$$

unter der Bedingung, dass das Management dem Vertrag zustimmt, d.h. die Behandlung eines Patienten der Fallschwere  $\theta$  zulässt. Die Anreizverträglichkeitsbedingung des Managements kommt mit der Reaktion des Managements auf die Kostenbeteiligung bei der Bestimmung der Budgetierung  $c^{d*}(\theta, r(\theta))$  zum Tragen.  $\theta B^*(.)$  stellt den Behandlungsnutzen für den Auftraggeber dar, wenn mit den aus der Beteiligung resultierenden Mengen  $q_1^*(c^{d*}(\theta, r(\theta)))$  und  $q_2^*(c^{d*}(\theta, r(\theta)))$  behandelt wird.

Analog zur Gleichung (42) lässt sich das Maximierungsproblem (72) auch äquivalent darstellen als

$$\max_{r(\theta), N\ddot{U}(\theta)} W = \theta B^*(.) + \theta V^*(.) - c^{d*}(\theta, r(\theta)) - N\ddot{U}(\theta) - u \quad (73)$$

unter der Nebenbedingung

$$N\ddot{U}(\theta) \geq g. \quad (74)$$

---

<sup>45</sup>Der Auftraggeber maximiert seinen Nutzen für jede Fallschwere  $\theta$ . Mit der Fallschwere als stetige Variable resultieren also „Pfade“ für die optimale Beteiligung  $r(\theta)$  und Pauschale  $k(\theta)$ . So gesehen entspricht das Problem des Auftraggebers einer punktwisen Maximierung.

Der gemeinsame Überschuss  $N\ddot{U}(\theta)$  von Arzt und Management geht negativ in die Zielfunktion des Auftraggebers ein. Im Optimum bindet also die Teilnahmebedingung des Managements (74) und für den gemeinsamen Nutzenüberschuss gilt

$$N\ddot{U}(\theta) = g. \quad (75)$$

Wir nehmen an, dass (73) streng konkav ist und erhalten mit (65) als Bedingung für die optimale Beteiligung bei einem Patienten der Fallschwere  $\theta$

$$\left[ \theta B_{q_1}^*(.) \frac{\partial q_1^*(.)}{\partial c^d(.)} + \theta B_{q_2}^*(.) \frac{\partial q_2^*(.)}{\partial c^d(.)} \right] \frac{\partial c^{d*}(.)}{\partial r(\theta)} + \lambda^*(\theta, .) \frac{\partial c^{d*}(.)}{\partial r(\theta)} - \frac{\partial c^{d*}(.)}{\partial r(\theta)} = 0, \quad (76)$$

wobei  $c^{d*}(.) = c^{d*}(\theta, r(\theta))$ . Der Term in den eckigen Klammern in Gleichung (76) ist die marginale Änderung des Behandlungsnutzens des Auftraggebers infolge einer marginalen Änderung der Budgetierung. Sie beinhaltet einerseits die Bewertung durch den Auftraggeber, andererseits die Reaktion des Arztes auf eine Budgetänderung. Multipliziert mit der aus einer marginalen Erhöhung der Beteiligung resultierenden Änderung der Budgetrestriktion durch das Management  $\frac{\partial c^{d*}(.)}{\partial r(\theta)}$  beziffert sie den marginalen Verlust des Behandlungsnutzens aus einer marginalen Erhöhung der Kostenbeteiligung. Der mittlere Term beschreibt die Änderung der (implizit gegebenen) Pauschale  $k(\theta)$ , die notwendig ist, um das Management auf dem Niveau seines Reservationsnutzens zu halten. Die Höhe  $\lambda^*(\theta, c^{d*}(.)) \frac{\partial c^{d*}(\theta, r(\theta))}{\partial r(\theta)}$  entspricht dem Betrag, mit dem das Management die freien Mittel für den Arzt anpasst (vgl. (65)). Der letzte Term beziffert die Änderung der Behandlungskosten. Im Optimum bestimmt der Auftraggeber die Beteiligung so, dass die Summe aus dem marginalen Nutzenverlust und der marginalen Erhöhung der Pauschale  $k(\theta)$  der marginalen Kostenersparnis entspricht.

Dividieren der Gleichung (76) durch  $\frac{\partial c^{d*}(.)}{\partial r(\theta)}$ <sup>46</sup> ergibt

$$\left[ \theta B_{q_1}^*(.) \frac{\partial q_1^*(.)}{\partial c^d(.)} + \theta B_{q_2}^*(.) \frac{\partial q_2^*(.)}{\partial c^d(.)} \right] + \lambda^*(\theta, .) - 1 = 0. \quad (77)$$

Mit impliziter Differenzierung folgt für das optimale Fallbudget

$$\frac{dc}{d\theta} = - \frac{\left[ B_{q_1}^*(.) \frac{\partial q_1^*(.)}{\partial c^d(.)} + B_{q_2}^*(.) \frac{\partial q_2^*(.)}{\partial c^d(.)} \right] + \partial \lambda^*(\theta, .) / \partial \theta}{\text{Bedingung zweiter Ordnung}} > 0. \quad (78)$$

Das Fallbudget zur Behandlung eines Patienten steigt also mit der Fallschwere. Für den optimalen Wert der Kostenbeteiligung folgt mit (66)

---

<sup>46</sup>Der Ausdruck  $\frac{\partial c^{d*}(\theta, r(\theta))}{\partial r(\theta)}$  ist mit (70) ungleich Null.



$$r^*(\theta) = 1 - \left[ \theta B_{q_1}^*(.) \frac{\partial q_1^*(.)}{\partial c^{d*}(.)} + \theta B_{q_2}^*(.) \frac{\partial q_2^*(.)}{\partial c^{d*}(.)} \right]. \quad (79)$$

Gleichung (79) zeigt, dass die optimale Kostenbeteiligung umso niedriger ist, je höher die marginale Änderung des Behandlungsnutzens des Auftraggebers infolge einer Budgetänderung ausfällt.

Die Pauschale  $k(\theta)$  wählt der Auftraggeber so, dass bei der optimalen Kostenbeteiligung die Teilnahmebedingung des Managements bindet und Arzt und Management zusammen den minimalen Nutzenüberschuss  $N\dot{U}(\theta) = g$  erhalten:

$$k^*(\theta) = r^*(\theta)c^{d*}(\theta, r^*(\theta)) - \theta V^*(.) + u + g. \quad (80)$$

Gleichungen (76) und (80) beschreiben in der Regel nicht die first best-Lösung (43), (44) und (45). Allerdings ist die first best-Lösung dann möglich, wenn im Gleichgewicht das Verhältnis der Grenznutzen des Arztes dem Verhältnis der Grenznutzen des Auftraggebers entspricht:

$$\frac{V_{q_1}^{FB}(.)}{V_{q_2}^{FB}(.)} = \frac{B_{q_1}^{FB}(.)}{B_{q_2}^{FB}(.)}. \quad (81)$$

Dies ist insbesondere dann der Fall, wenn die ärztliche Bewertung der Behandlung eine lineare Funktion der Bewertungsfunktion des Auftraggebers ist. Wenn die Bedingung (81) erfüllt ist, dann lassen sich die Bedingungen erster Ordnung aus dem Maximierungsprogramm des Arztes (48) und (49) durch geeignete Wahl der Zahlungsbereitschaft  $\lambda^*(\theta, c^d(\theta))$ , d.h. durch geeignete Wahl der Budgetierung  $c^{d*}(\theta)$  in die Bedingungen erster Ordnung für eine first best-Lösung (43) und (44) überleiten.

In diesem Falle sind Auftraggeber und Arzt über das Verhältnis der Leistungsmengen am Punkt der first best-Lösung der gleichen Meinung. Der Auftraggeber muss dann nur mit der geeigneten Wahl der Kostenbeteiligung  $r(\theta)$  dem Management Anreize geben, dem Arzt das für die first best-Mengen  $q_1^{FB}(\theta)$  und  $q_2^{FB}(\theta)$  notwendige Budget vorzugeben. Die Zusammensetzung der Leistungen wählt der Arzt dann im Sinne des Auftraggebers. Das zweidimensionale Moral-Hazard-Problem reduziert sich dann also zu einem eindimensionalen Problem, bei dem der Auftraggeber dem Management lediglich den Anreiz setzen muss, die richtige spitalinterne Budgetierung zu bestimmen.

Wenn eine first best-Lösung möglich ist, dann lässt sich der optimale Grad der Kostenbeteiligung alternativ zu Gleichung (79) auch aus den Gleichungen (43), (44), (48), (49)

und der Bedingung erster Ordnung für die optimale Budgetierung durch das Management (66) für die optimale Kostenbeteiligung berechnen:

$$r^{FB}(\theta) = \frac{V_{q_1}^{FB}(\cdot)}{V_{q_1}^{FB}(\cdot) + B_{q_1}^{FB}(\cdot)} \left( = \frac{V_{q_2}^{FB}(\cdot)}{V_{q_2}^{FB}(\cdot) + B_{q_2}^{FB}(\cdot)} \right). \quad (82)$$

Da bei der first best-Lösung  $\theta V_{q_1} > 0$  und  $\theta B_{q_1} > 0$  (bzw.  $\theta V_{q_1} > 0$  und  $\theta B_{q_1} > 0$ ) gilt, liegt die optimale Kostenbeteiligung im Bereich  $(0, 1)$ . Sind die Bewertungen identisch, resultiert eine Kostenbeteiligung von 0.5.<sup>47</sup> Bewertet der Arzt die marginale Ausweitung der Mengen höher als der Auftraggeber, ist die Kostenbeteiligung, die zur first best-Lösung führt, grösser als 0.5. Der Auftraggeber muss in diesem Fall dem Management relativ starke Anreize zu geben, die Budgetierung niedrig zu halten, damit der Arzt nicht zu hohe Mengen einsetzt. Umgekehrt liegt die Kostenbeteiligung unter 0.5, wenn die marginale Bewertung durch den Arzt niedriger ist als die des Auftraggebers.

Unterschiede zwischen den marginalen Bewertungen von Arzt und Auftraggeber (und auch in den Sättigungsmengen) sind vor allem dann zu erwarten, wenn beim Arzt auch egoistische Motivationen bei der Behandlung eine Rolle spielen. Man denke an besonders prestigeträchtige Behandlungsmethoden, die besondere Fähigkeiten des Arztes voraussetzen oder an Zuwendungen, die ein Arzt für die Verwendung eines bestimmten Produktes vom Hersteller erhält. In diesem Fall dürfte die marginale Bewertung und evtl. die Sättigungsmenge des Arztes über jener des Auftraggebers liegen. Auf der anderen Seite dürften beispielsweise Routineleistungen beim Arzt im Vergleich zum Auftraggeber keine allzu hohe Wertschätzung erfahren, so dass bei diesen Leistungen die marginale Bewertung unter jener des Auftraggebers liegen wird.

Wir fassen die wesentlichen Erkenntnisse aus diesem Abschnitt zusammen:

**Ergebnis 6.1** *Bei spitalinterner PA-Lösung berücksichtigt das Management bei der Budgetierung vollständig die Präferenzen des Arztes hinsichtlich der Mengen  $q_1$  und  $q_2$ , so dass aus Sicht des Auftraggebers das Spital ein Eigeninteresse an der Behandlungsqualität hat, selbst wenn das Management indifferent gegenüber der Qualität der Behandlung ist. Daher kann der Auftraggeber dem Management einen Teil der Kostenverantwortung übertragen, ohne dass spitalintern ein minimales Budget vom Management bestimmt wird, welches gerade die minimalen Behandlungsmengen ermöglicht. Eine first best-Lösung ist für den Auftraggeber aber in der Regel nicht zu erreichen, da weder er noch das Management die vom Arzt gewählte Zusammensetzung der Leistungen bestimmen kann.*

<sup>47</sup>In diesem Zusammenhang sprechen Ellis und McGuire von einem „perfekten Agenten“, wenn sich die Bewertungsfunktionen von Arzt und Management lediglich durch einen konstanten Betrag unterscheiden [Ellis/McGuire (1986)], d.h. wenn für eine Konstante  $s$  gilt:  $\theta V(q_1, q_2) = \theta B(q_1, q_2) + s$ .

### 6.3.7 Bedeutungslosigkeit der spitalinternen Informationsasymmetrie

Es zeigt sich, dass es im Modell für die resultierende Allokation keine Rolle spielt, ob das Management die Quantitäten  $q_1$  und  $q_2$  beobachten und dem Arzt vorschreiben, oder lediglich die Einhaltung der Budgetierung fordern kann und die Zusammensetzung der Mengen dem Arzt überlassen muss. Das Management würde exakt die gleichen Quantitäten wählen, die der Arzt unter Vorgabe des optimalen Budgets  $c^{d*}(\theta, r(\theta))$  einsetzt.

Angenommen, die Mengen seien für das Management beobachtbar und kontrahierbar, dann ergibt sich für das Maximierungsproblem des Managements

$$\max_{q_1(\theta), q_2(\theta)} G = k(\theta) - r(\theta)(p_1 q_1(\theta) + p_2 q_2(\theta)) + \theta V(q_1(\theta), q_2(\theta)) - u. \quad (83)$$

Als Bedingungen erster Ordnung für die optimalen Mengen aus Sicht des Managements folgen

$$\theta V_{q_1}(q_1^*(\theta), q_2^*(\theta)) = r(\theta)p_1 \quad (84)$$

$$\theta V_{q_2}(q_1^*(\theta), q_2^*(\theta)) = r(\theta)p_2. \quad (85)$$

Nun entsprechen die Bedingungen erster Ordnung (84) und (85) exakt den Bedingungen (48) und (49), die sich aus dem Maximierungsproblem des Arztes ergeben, wenn dieser die Mengen unter Einhaltung der Budgetierung  $c^{d*}(\theta, r(\theta))$  wählt, für welche die Bedingung (66),  $\lambda^*(\cdot) = r(\theta)$  gilt.

Der Grund ist einleuchtend: Das Management sucht den optimalen Trade-off zwischen möglichst niedrigen Behandlungskosten und möglichst niedrigen freien Mitteln für den Arzt. Je restriktiver die Budgetierung, umso niedriger ist der Nutzen  $\theta V(q_1, q_2)$ , den der Arzt aus der Behandlung zieht und umso höher müssen die notwendigen freien Mittel sein, um dem Arzt seinen Reservationsnutzen zu gewähren. Da es keine eigenen Präferenzen gegenüber den Mengen  $q_1$  und  $q_2$  hat, wäre es für das Management optimal, bei gegebenem Budget die Mengenkombination zu wählen, die den Behandlungsnutzen des Arztes maximiert. Die Lösung ist aber genau die Lösung des Maximierungsproblems des Arztes aus Abschnitt 6.3.2. Daher spielt es keine Rolle, ob das Management die Mengen dem Arzt vorschreiben kann oder lediglich eine Budgetvorgabe definiert.

Für den Auftraggeber hat die Irrelevanz der internen asymmetrischen Information bezüglich der Mengen  $q_1$  und  $q_2$  insofern Bedeutung, als er die genauen Umstände der Informationsverteilung im Spital nicht kennen muss, wenn er die optimalen Vergütungsparameter

$r(\theta)$  und  $k(\theta)$  bestimmen will.

**Ergebnis 6.2** *Hat das Management keine eigenen Präferenzen bezüglich der Behandlungsmengen, muss der Auftraggeber bei der Wahl der Vergütungsparameter  $k(\theta)$  und  $r(\theta)$  die Informationsasymmetrie zwischen Arzt und Management hinsichtlich  $q_1$  und  $q_2$  nicht beachten.*

## 6.4 Lösung des Modells bei spitalinterner Verhandlungslösung

Nun stellt sich die Frage, welchen Bestand die Ergebnisse des vorigen Abschnittes haben, wenn der interne Prinzipal-Agent-Mechanismus mit dem vom Management definierten take-it-or-leave-it Angebot durch eine Verhandlungslösung zwischen Management und Arzt ersetzt wird und der Arzt durch Verhandlungsmacht und Verhandlungsgeschick versucht, seine Interessen einzubringen. Wir gehen dieser Frage nach, in dem wir zur theoretischen Darstellung der Verhandlung die Parameter  $t^d(\theta)$  und  $c^d(\theta)$  als das Ergebnis einer Nash-Verhandlungslösung bestimmen.

Die Nash-Verhandlungslösung ist der spieltheoretische Standardansatz, um das Ergebnis von Verhandlungen zu modellieren [vgl. Barros (2006)]. Dieser Ansatz hat neben einer eingängigen Interpretation den Vorteil, dass so etwas wie Verhandlungsmacht in die Modellierung einbezogen werden kann.

### 6.4.1 Zeitliche Abfolge

Wir nehmen an, dass die Verhandlungen nach dem Angebot des Auftraggebers, jedoch vor der Behandlung des ersten Patienten durch den Arzt stattfinden, so dass sich folgender zeitlicher Ablauf ergibt:

1. In der ersten Stufe des Spieles bestimmt der Auftraggeber die Parameter  $k(\theta)$  und  $r(\theta)$  für jede Fallschwere  $\theta \in \Theta$ .
2. In der zweiten Stufe entscheidet das Management über die Annahme oder die Ablehnung des Vertragsangebotes des Auftraggebers. Wenn es das Angebot annimmt, verhandeln das Management und der Arzt für jede Fallschwere  $\theta \in \Theta$  das Budget  $c^d(\theta)$  und die freien Mittel  $t^d(\theta)$ .
3. In der dritten Stufe kommen die  $n$  Patienten zur Behandlung in das Spital. Der Arzt stellt die jeweilige Fallschwere  $\theta$  der einzelnen Patienten fest und entscheidet, ob er den Patienten zur Behandlung aufnimmt oder ob er ihm die Aufnahme verweigert. Wenn er sich für die Behandlung des Patienten entscheidet, berichtet der Arzt

die Fallschwere wahrheitsgemäss an das Management. Das Management entscheidet, ob es die Behandlung zulässt. Wenn das Management die Behandlung zulässt, bestimmt der Arzt die Mengen  $q_1, q_2$  für die Behandlung des Patienten so, dass das zwischen Arzt und Management verhandelte Budget  $c^d(\theta)$  eingehalten wird. Das Management berichtet die Fallschwere wahrheitsgemäss an den Auftraggeber. Die Zahlungen finden statt, insbesondere die der verhandelten freien Mittel  $t^d(\theta)$ .

Wie im vorigen Abschnitt wird das Spiels mittels Rückwärtsinduktion gelöst.

#### **6.4.2 Bestimmung der Mengen $q_1$ und $q_2$ durch den Arzt**

In der letzten Stufe des Spieles maximiert der Arzt seinen Nutzen mit der Bestimmung der Mengen  $q_1(\theta)$  und  $q_2(\theta)$ , gegeben dem in der Verhandlung festgelegten Budget  $c^d(\theta)$  und den exogenen Preisen  $p_1$  und  $p_2$ . Wir nehmen an, das Management verfügt über ausreichende Sanktionsmöglichkeiten, um zu verhindern, dass der Arzt das verhandelte Budget nicht einhält. Das Maximierungsproblem ist damit identisch mit jenem in Abschnitt 6.3.2 und es resultieren dieselben Ergebnisse.

#### **6.4.3 Bestimmung der Budgetierung $c^d(\theta)$ sowie der freien Mittel $t^d(\theta)$ durch Verhandlungen zwischen Management und Arzt**

In der zweiten Stufe legen das Management und der Arzt im Rahmen von Verhandlungen die Parameter  $t^d(\theta)$  und  $c^d(\theta)$  fest. Wie zu Beginn dieses Abschnittes erwähnt, setzen wir die theoretische Darstellung der Verhandlung mit der Nash-Verhandlungslösung um.

Der stilisierte Verhandlungsprozess zwischen Management und Arzt besteht aus einer Abfolge aus Angebot und Gegenangebot und dauert so lange, bis ein Verhandlungspartner auf das aktuelle Angebot  $\{t^d(\theta), c^d(\theta)\}$  des anderen Verhandlungspartners eingeht. Zwischen den Angeboten der beiden Verhandlungspartner liegt immer ein gewisser zeitlicher Abstand, so dass vom Zeitpunkt des Verhandlungsbeginns aus gesehen das Verhandlungsergebnis in der Zukunft liegt. Nun diskontieren Arzt und Management Auszahlungen, die in der Zukunft liegen, ab, zum Beispiel weil die Opportunitätskosten steigen, je länger die Verhandlung dauert, oder weil die Verhandlung an sich finanzielle und nicht-finanzielle Kosten verursacht.<sup>48</sup>

Wenn man so will, lässt sich aus dem Ausmass, mit dem der Arzt im Vergleich zum Management eine zukünftige Auszahlung abdiskontiert, seine Verhandlungsmacht gegenüber dem Management ableiten. Wenn  $w \in [0, 1]$  die Verhandlungsmacht des Managements

---

<sup>48</sup>Dieser Verhandlungsprozess ist als Rubinstein-Verhandlungsmodell bekannt [vgl. Muthoo (1999), Kapitel 3].

gegenüber dem Arzt darstellt, bzw.  $1 - w$  die Verhandlungsmacht des Arztes gegenüber dem Management, dann gilt mit  $\rho^d$  als Diskontfaktor des Arztes und  $\rho^m$  als jenem des Managements  $w = \frac{\rho^d}{\rho^d + \rho^m}$  und  $1 - w = \frac{\rho^m}{\rho^d + \rho^m}$ , d.h. die Verhandlungsmacht des Managements gegenüber dem Arzt ist umso grösser, je grösser der Diskontfaktor des Arztes gegenüber dem des Managements ist und umgekehrt.<sup>49</sup>

Allerdings darf die Verteilung der Verhandlungsmacht nicht dazu führen, dass eine Lösung resultiert, die für einen Verhandlungspartner die Unterschreitung seines internen Reservationsnutzens bedeutet. Zeichnet sich im Laufe der Verhandlungen ab, dass für den Arzt ein Nutzen resultiert, der unter dem Reservationsnutzen  $u$  liegt, bricht der Arzt die Verhandlungen ab und andere Mechanismen greifen, die zum Reservationsnutzen  $u$  führen. Gleiches gilt auch für das Management, welches die Verhandlung abbricht, wenn nicht der Mindestgewinn  $a$  gewährleistet ist.

Im Verlauf der Verhandlungen wird der Arzt also durch zwei Parameter charakterisiert: Zum einen durch seinen Reservationsnutzen, der sich als Stärke des Chefarztes interpretieren lässt, die er in die Verhandlung mitbringt und die er aus Eigenschaften wie Kompetenz und medizinischen Ruf zieht. Je mehr der Arzt eine Grösse auf seinem Gebiet darstellt, umso hochwertiger werden seine Alternativen sein. Zum anderen spiegelt  $1 - w$  den Willen und die Fähigkeit des Arztes wieder, mehr aus den Verhandlungen zu holen als nur den Reservationsnutzen. Dieser Parameter hat mit der „Geduld“ des Arztes im Vergleich zum Management zu tun.

Es lässt sich zeigen, dass der dargestellte Verhandlungsprozess unter gewissen Voraussetzungen, die hier erfüllt sind,<sup>50</sup> ein eindeutiges Gleichgewicht  $\{t^{d**}(\theta), c^{d**}(\theta)\}$  besitzt, das gegen die Nash-Verhandlungslösung  $\{t^{d*}(\theta), c^{d*}(\theta)\}$  konvergiert, wenn der Zeitraum zwischen Angebot und Gegenangebot gegen Null geht. Das kommt einer Situation nahe, in der die Verhandlung in einer Sitzung abgeschlossen wird.<sup>51</sup>

---

<sup>49</sup>Intuitiv stellt sich die Frage, welche Bedeutung ein Diskontfaktor für zukünftige Auszahlungen hat, wenn der Zeitraum zwischen den Angeboten gegen null geht. Hier ist zu beachten, dass selbst wenn die Zeit zwischen Angebot und Gegenangebot und damit die Opportunitätskosten gegen null gehen, die Relation der Diskontfaktoren von Arzt und Management, die letztendlich für die Aufteilung des gemeinsamen Überschusses verantwortlich ist, weiterhin besteht [Muthoo [1999, Seiten 51-52].

<sup>50</sup>Die Voraussetzungen sind zum einen, dass die Pareto-effiziente Grenze konkav ist, und sich zu jedem Paar  $\{G(\cdot), U(\cdot)\}$ , das auf der Pareto-effizienten Grenze liegt, genau eine Kombination  $\{t^d(\theta), c^d(\theta)\}$  zuordnen lässt [Muthoo (1999), Seiten 59 - 65].

<sup>51</sup>Eine ausführliche Darstellung des Rubinstein-Verhandlungsmodells sowie ein umfangreicher Beweis der Existenz eines eindeutigen Gleichgewichtes und die Konvergenz desselben zur Nash-Verhandlungslösung ist in Muthoo zu finden (Muthoo 1999, Kap. 3).

Die Nash-Verhandlungslösung ihrerseits ist die Lösung des Maximierungsproblems

$$\max_{c^d(\theta), t^d(\theta)} \pi(\theta) = (G(\theta) - a)^w (U(\theta) - u)^{1-w} \quad (86)$$

wobei

$$G(\theta) = k(\theta) - r(\theta)c^d(\theta) - t^d(\theta)$$

und

$$U(\theta) = t^d(\theta) + \theta V^*(.).$$

In (86) wird zur Veranschaulichung der Mindestgewinn für die interne Teilnahme des Managements  $a = 0$  notiert. Das Verhandlungsergebnis muss für den Arzt mindestens den Reservationsnutzen  $u$  bzw. für das Management einen Mindestgewinn  $a$  mit sich bringen. Hier wird die Bedeutung der externen Teilnahmebedingung  $N\ddot{U}(\theta) \geq g > a = 0$  deutlich, die sicherstellt, dass es zwischen Management und Arzt überhaupt einen Nutzenüberschuss zum Verhandeln gibt, auch wenn der Auftraggeber mit seinem Vertrag dem Management gerade einen Nutzenüberschuss in Höhe von  $g$  zugesteht.

Rechtfertigen lässt sich die Annahme mit der Überlegung, dass zum Zeitpunkt der Verhandlung das Management bereits dem Vertrag mit dem Auftraggeber zugesagt hat und von daher ein Scheitern der internen Verhandlungen vermeiden will, um eventuelle Konventionalstrafen infolge von Nicht-Erfüllung zu vermeiden.

Die Bedingungen erster Ordnung für die verhandelten optimalen freien Mittel  $t^{d*}(\theta)$  für ein Maximum von (86) ergibt (mit  $a = 0$ )

$$\frac{w}{G(\theta)} = \frac{1-w}{U(\theta) - u}. \quad (87)$$

Die freien Mittel werden also derart verhandelt, dass der Verhandlungspartner mit der grösseren Verhandlungsmacht den grösseren Nutzenüberschuss erhält. Die Bedingung erster Ordnung für die verhandelte Budgetierung  $c^{d*}(\theta, r(\theta))$  lautet

$$-\frac{w}{G(\theta)}r(\theta) + \frac{1-w}{U(\theta) - u}\lambda^*(\theta, .) = 0. \quad (88)$$

Mit der Verteilung des Überschusses (87) ergibt sich für (144)

$$\lambda^*(\theta, \cdot) - r(\theta) = 0. \quad (89)$$

Gleichung (89) entspricht der Bedingung erster Ordnung (66) für die optimale Budgetierung der PA-Lösung in Abschnitt 6.3.4. Das Verhandlungsergebnis liefert also dasselbe Fallbudget und damit dieselben Behandlungsmengen wie die PA-Lösung. Weder die Verhandlungsmacht  $1 - w$  noch der Reservationsnutzen  $u$  des Arztes spielt für das optimale verhandelte Budget eine Rolle.

Die Verteilung des Nutzenüberschusses nach der Regel (87) erfolgt also ausschliesslich über die freien Mittel, d.h. je grösser die Verhandlungsmacht des Arztes  $1 - w$ , umso höher sind die verhandelten freien Mittel  $t^{d*}(\theta)$ .

Mit den Bedingungen (87) und (89) lässt sich das optimale Budget als eine Funktion der Kostenbeteiligung darstellen, die optimalen freien Mittel sind dagegen eine Funktion der Kostenbeteiligung *und* der internen Machtverhältnisse:

$$c^{d*}(\theta) \equiv c^{d*}(\theta, r(\theta)). \quad (90)$$

$$t^{*d}(\theta) \equiv t^{*d}(\theta, r(\theta), w). \quad (91)$$

Der Grund für die Unabhängigkeit der optimalen Budgetierung von der Verteilung der Verhandlungsmacht und der entsprechenden Verteilung des Nutzenüberschusses ist in der Risikoneutralität des Managements und des Arztes gegenüber der freien Mittel zu sehen. Wegen der Risikoneutralität ist es möglich, mit den freien Mitteln den gemeinsamen Nutzenüberschuss von Arzt und Management zwischen den beiden nach der Regel (87) zu verteilen, ohne dass sich der Betrag des gemeinsamen Nutzenüberschuss dadurch ändert. Daher macht es Sinn, wenn sich Arzt und Management in der Verhandlung auf eine Budgetierung einigen, die den gemeinsamen Nutzenüberschuss maximiert, um dann den Überschuss gemäss den Verhältnissen der Verhandlungsmacht spitalintern untereinander aufzuteilen.

Abbildung 2 verdeutlicht dies. Auf der horizontalen Achse ist der Gewinn des Managements bei der Behandlung eines Patienten der Fallschwere  $\theta$  abgetragen, auf der vertikalen der Nutzen des Arztes (die Fallschwere ist in der Darstellung unterdrückt). Die Gerade durch die Punkte NV und PA stellt die Pareto-effiziente Grenze dar, sie hat die Steigung  $dG/dU = -1$ . Per Definition gibt sie für jedes Nutzenniveau  $\bar{U}$  des Arztes den maximal



erreichbaren Gewinn des Managements an. Da es für den Auftraggeber optimal ist, dem Spital (Arzt und Management) durch die Wahl der Pauschale  $k$  gerade seinen Reservationsüberschuss  $g$  zu gewähren, muss die Grenze durch die Punkte  $(g, u)$  und  $(0, u + g)$  verlaufen.

Die Verhandlungslösung (Punkt NV) liegt auf der Pareto-effizienten Grenze, da die gemeinsame Zielfunktion (86), die von der Verhandlungslösung maximiert wird, monoton im Nutzen des Arztes und im Gewinn des Managements steigt. Lösungen links der Grenze sind nicht optimal, da dann der Gewinn des Managements bei konstantem Nutzen des Arztes erhöht werden könnte. Die Isoquanten der gemeinsamen Zielfunktion (86) sind konvex zum Ursprung. Je weiter rechts eine Isoquante liegt, umso höher ist ihr Wert. Die Isoquante, zu der die Pareto-effiziente Grenze tangential ist, enthält den maximal möglichen Wert, da Lösungen rechts der Pareto-effizienten Grenze nicht möglich sind. Das Verhandlungsergebnis liegt also im Tangentialpunkt NV. Je kleiner die Verhandlungsmacht  $w$  des Managements, desto weiter links oben auf der Pareto-effizienten Grenze liegt der Tangentialpunkt. Im dargestellten Beispiel beschert die Lösung dem Arzt einen Nutzen in Höhe von  $\tilde{U}^* > u$  und dem Management einen Gewinn in Höhe von  $\tilde{G}^* < g$ .

Die PA-Lösung (Punkt PA) liegt ebenfalls auf der Pareto-effizienten Grenze, da sie den Gewinn des Managements bei gegebenem Reservationsnutzen  $u$  des Arztes maximiert. Sie enthält für den Arzt den Reservationsnutzen  $u$ , während das Management den Gewinn  $g$  und damit den gesamten Nutzenüberschuss erhält. Da die beiden Lösungen Pareto-effizient sind, erfüllen sie die Bedingung  $\lambda^*(\theta, c^{d*}(\cdot)) = r(\theta)$  und führen damit zur gleichen optimalen Budgetierung  $c^{d*}(\theta)$ . Beide Lösungen unterscheiden sich aber in den freien Mitteln. Im Verhandlungsfall einigen sich Arzt und Management auf freie Mittel in Höhe von  $t^{d*}(\theta) = \tilde{G}(\theta) - V^*(\cdot)$ , in der PA-Lösung legt das Management die freien Mittel in Höhe von  $t^{d*}(\theta) = u - V^*(\cdot)$  fest.

Für die Nutzenmaximierung des Auftraggebers ändert sich durch die Verhandlungslösung also nichts. Für eine gegebene Kostenverantwortung  $r(\theta)$  folgt dasselbe Budget  $c^{d*}(\theta, r(\theta))$  und damit die gleichen Mengen  $q_1^*(\cdot)$  und  $q_1^*(\cdot)$ . Aus Sicht des Auftraggebers reagiert das Spital exakt gleich auf die durch die Kostenbeteiligung gesetzten Anreize, und es ist weiterhin optimal, wenn der Auftraggeber die Pauschale so wählt, dass das Spital gerade den Nutzenüberschuss  $g$  erhält.

**Ergebnis 6.3** *Wenn der Arzt und das Management risikoneutral gegenüber den freien Mitteln sind, dann muss der Auftraggeber für die Bestimmung der aus seiner Sicht optimalen Kostenbeteiligung  $r(\theta)$  und der Pauschale  $k(\theta)$  nicht über den internen Mecha-*

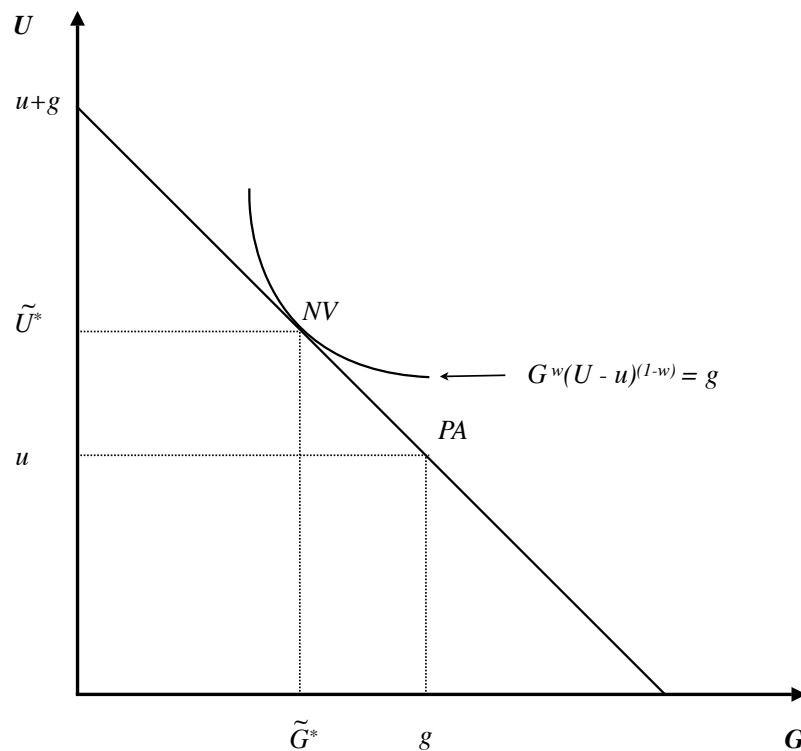


Abbildung 2: Äquivalenz von Verhandlungslösung und PA-Lösung aus Sicht des Auftraggebers. Beide Lösungen unterscheiden sich in der Höhe der freien Mittel und damit in der Verteilung des Nutzenüberschusses, nicht aber im optimalen Budget. Die Verhandlungslösung konvergiert zur PA-Lösung, wenn die Verhandlungsmacht des Managements vollständig beim Management liegt, d.h. wenn  $w \rightarrow 1$ .

nismus informiert sein. Für den Auftraggeber ist es gleichgültig, ob das Management und der Arzt spitalintern in einem Prinzipal-Agent-Verhältnis zueinander stehen und das Management die Budgetierung und die freien Mittel festlegt, oder ob das Budget und die freien Mittel das Ergebnis einer Nash-Verhandlungslösung sind. Insbesondere sind dabei die Machtverhältnisse zwischen Arzt und Management ohne Bedeutung.

## 6.5 Risikoaversion des Arztes gegenüber der freien Mittel

Wie wir gesehen haben, spielt es aus Sicht des Auftraggebers keine Rolle, ob die interne Fallbudgetierung  $c^d(\theta)$  und die freien Mittel für den Arzt  $t^d(\theta)$  bei einem spitalinternen PA-Verhältnis vom Management festgelegt werden oder das Ergebnis von Verhandlungen zwischen Management und Arzt sind, wenn der gemeinsame Nutzenüberschuss über die freien Mittel zwischen Arzt und Management aufgeteilt werden kann, ohne dabei den gemeinsamen Nutzenüberschuss in seiner Höhe zu beeinflussen. Das ist möglich, wenn Management und Arzt risikoneutral gegenüber den freien Mitteln sind (und die freien Mittel keine „Schattenkosten“ verursachen).

Ist dagegen zum Beispiel der Arzt den freien Mitteln gegenüber risikoavers eingestellt, führt eine über die freien Mittel finanzierte Verschiebung des Anteils des gemeinsamen Nutzenüberschusses zugunsten des Arztes zu einer Reduktion des gemeinsamen Überschusses, da der Arzt jede zusätzliche Einheit an Geld geringer bewertet. Die Vermutung liegt nahe, dass in diesem Fall die Äquivalenz der PA-Lösung und der NV-Lösung nicht mehr gegeben ist (obwohl beide Lösungen zu einer Pareto-effizienten Allokation zwischen Management und Arzt führen).

Um die Analyse auf den Effekt der Risikoaversion des Arztes gegenüber den freien Mitteln zu konzentrieren, nehmen wir an, der Nutzen des Arztes aus den freien Mitteln sei konkav. Wir halten jedoch an der Annahme fest, dass die Nutzenfunktion des Arztes additiv-separabel in den freien Mitteln und der Bewertung der Behandlung ist. Wir postulieren für die Nutzenfunktion des Arztes:

$$U(\theta) = H(t^d(\theta)) + \theta V(.), \quad (92)$$

wobei  $H(t^d)$  die Bewertung der freien Mittel ist, mit  $H'(t^d) > 0$ ,  $H''(t^d) < 0$ . Ausserdem machen wir die plausiblen Annahmen, dass der Arzt zum einen freie Mittel von null auch mit einem Wert von null bewertet und zum anderen der marginale Nutzen aus den freien Mitteln immer unterhalb 1 liegt, also unter dem marginalen Nutzen des risikoneutralen Arztes. Es gilt also  $H(0) = 0$  und  $H'(0) \leq 1$ , was wiederum  $H'(t^d) < 1 \forall t^d > 0$  impliziert.

Da  $H(t^d)$  streng konkav ist, lassen sich die freien Mittel  $t^d(\theta)$  mit der Umkehrfunktion

$$t^d(\theta) = H^{-1}(U(\theta) - \theta V(.)) \quad (93)$$

eindeutig als Funktion der Differenz zwischen dem Gesamtnutzen  $U(\theta)$  und der Bewertung  $\theta V(.)$  darstellen. Die Umkehrfunktion ist streng konvex und positiv, wenn  $U(\theta) - \theta V(.)$  positiv ist. Sie hat im relevanten Bereich  $U(\theta) - \theta V(.) > 0$  wegen  $H^{-1'}(.) = \frac{1}{H'(t^d)} > 1$  überall eine Steigung grösser eins [vgl. Chiang (1984)].

### 6.5.1 First best-Lösung

Einsetzen von (93) in die Zielfunktion (62) des Managements und weiteres Einsetzen in die Zielfunktion des Auftraggebers (40) ergibt für das Maximierungsproblem des Auftraggebers bei der Behandlung eines Patienten der Fallschwere  $\theta$  und bei vollständiger Information

$$\max_{q_1(\theta), q_2(\theta), U(\theta), NU(\theta)} W(\theta) = \theta B(.) - p_1 q_2(\theta) - p_2 q_2(\theta) - NU(\theta) - H^{-1}(U(\theta) - \theta V(.)) + U(\theta) - u. \quad (94)$$

Die beiden jeweiligen Bedingungen erster Ordnung für die first best-Mengen  $q_1^{FB}(\theta)$  und  $q_2^{FB}(\theta)$  bei der Behandlung eines Patienten der Fallschwere  $\theta$  lauten:

$$\theta B_{q_1}(q_1^{FB}(\theta), q_2^{FB}(\theta)) + H^{-1'}(.) \theta V_{q_1}(q_1^{FB}(\theta), q_2^{FB}(\theta)) = p_1 \quad (95)$$

$$\theta B_{q_2}(q_1^{FB}(\theta), q_2^{FB}(\theta)) + H^{-1'}(.) \theta V_{q_2}(q_1^{FB}(\theta), q_2^{FB}(\theta)) = p_2 \quad (96)$$

Die Bedingungen (95) und (96) unterscheiden sich von den Bedingungen erster Ordnung im Falle eines risikoneutralen Arztes (43) und (44) zum einen dadurch, dass der Auftraggeber die marginale Bewertung des Arztes mit  $H^{-1'}(.)$  gewichtet, um dessen Risikoaversion zu berücksichtigen. Da  $H^{-1'}(.) > 1$ , spricht der Auftraggeber der Bewertung ein höheres Gewicht zu als im Falle eines risikoneutralen Arztes.

Ein weiterer Unterschied ergibt sich hinsichtlich des Nutzens  $U(\theta)$  bzw. Gewinns  $G(\theta)$ , den Arzt bzw. Management bei der first best-Lösung erhalten. Wegen  $H^{-1'}(.) > 1$  geht neben dem Nutzenüberschuss auch explizit der Nutzen des Arztes in die Zielfunktion eindeutig negativ ein, so dass im Optimum

$$N\ddot{U}(\theta) = G(\theta) = g \quad (97)$$

$$U(\theta) = u \quad (98)$$

erfüllt sein muss.

**Ergebnis 6.4** *Ein risikoaverser Arzt erhält bei der first best-Lösung gerade seinen Reservationsnutzen, das Management erhält dagegen den gesamten spitalinternen Nutzenüberschuss.*

Im Vergleich ist im Falle des risikoneutralen Arztes für die first best-Lösung hinsichtlich der Nutzen von Arzt und Management lediglich entscheidend, dass der gemeinsame Überschuss den tiefstmöglichen Wert  $g$  annimmt. Die konkrete Verteilung des Überschusses zwischen Arzt und Management ist für die first best-Lösung bei einem risikoneutralen Arzt nicht relevant. Damit wird bereits hier klar, dass eine first best-Lösung nicht zu erreichen ist, wenn der Arzt bei spitalinterner NV-Lösung einen Nutzen oberhalb seines Reservationsnutzens erhält.

## 6.6 Lösung des Modells bei spitalinternem PA-Verhältnis und Risikoaversion des Arztes

### 6.6.1 Bestimmung der Mengen $q_1$ und $q_2$ durch den Arzt

Durch die additive Separabilität der Zielfunktion des Arztes stellen wir sicher, dass der Arzt seine Entscheidung über die zur Behandlung eingesetzten Mengen unabhängig von den freien Mitteln fällt. Die Bedingungen für die aus Sicht des Arztes optimalen Behandlungsmengen bei gegebenen Budget,  $q_1^*(.)$  und  $q_2^*(.)$ , sowie die Zahlungsbereitschaft  $\lambda^*(\theta, .)$  bleiben von der Risikoaversion gegenüber den freien Mitteln unberührt und sind damit auch hier gültig.

### 6.6.2 Bestimmung der Budgetierung $c^d(\theta)$ sowie der freien Mittel $t^d(\theta)$ durch das Management

Mit der Bestimmung der Mengen durch den Arzt ergibt für (93)

$$t^d(\theta) = H^{-1}(U(\theta) - \theta V^*(.)). \quad (99)$$

Einsetzen von (99) in seine Zielfunktion (62) ergibt das Maximierungsproblem des Managements, wenn es mit einem risikoaversen Arzt zu tun hat:

$$\max_{c^d(\theta), t^d(\theta)} G(\theta) = k(\theta) - r(\theta)c^d(\theta) - H^{-1}(U(\theta) - \theta V^*(.)). \quad (100)$$

Aus  $H^{-1'}(.) > 0$  folgt, dass es für das Management optimal ist, wenn der Arzt auch dann gerade seinen Reservationsnutzen erhält, wenn er risikoavers ist.<sup>52</sup> Im Optimum gestattet das Management dem Arzt also gerade dessen Reservationsnutzen zu

$$U(\theta) = u. \quad (101)$$

Die Bedingung erster Ordnung für die optimale Budgetierung hingegen lautet im Optimum

$$-r(\theta) + H^{-1'}(u - \theta V^*(.))\lambda^*(\theta, c^{d*}(\theta)) = 0. \quad (102)$$

und ist, im Gegensatz zu derjenigen im Falle eines risikoneutralen Arztes, abhängig von der Höhe der freien Mittel ( $t^d(\theta) = H^{-1}(u - \theta V^*(.))$ ). Die Bedingung zweiter Ordnung

$$-H^{-1''}(u - \theta V^*(.))\lambda^*(\theta, .)^2 + H^{-1'}(u - \theta V^*(.))\frac{\partial \lambda^*(\theta, .)}{\partial c^d(\theta)}. \quad (103)$$

ist erfüllt, da die Umkehrfunktion  $t^d(\theta) = H^{-1}(u - \theta V^*(.))$  streng konvex ist und wir  $\frac{\partial \lambda^*(\theta, .)}{\partial c^d(\theta)} < 0$  annehmen.

Aus Gleichung (102) geht hervor, dass im Falle des risikoaversen Arztes das aus Sicht des Managements optimale Budget mit einer Zahlungsbereitschaft für eine Budgetausweitung  $\lambda^*(.)$  verbunden ist, die kleiner ist als die Kostenbeteiligung,  $\lambda^*(.) < r(\theta)$ . Bei gegebener Kostenbeteiligung wählt das Management also bei einem risikoaversen Arzt eine höhere Budgetierung als bei einem risikoneutralen Arzt. Der Grund ist, dass sich durch die Risikoaversion der Trade-off zwischen freien Mitteln und Budgetierung mehr zugunsten der Budgetierung verschiebt, da der marginale Nutzen des Arztes aus den freien Mitteln kleiner eins ist, so dass das Management im Falle des risikoaversen Arztes eine Lösung sucht, die mehr Gewicht auf die Budgetierung legt. Dies äussert sich auch in einer höheren Sensitivität des Managements gegenüber den vom Auftraggeber gesetzten Anreizen. Totales Differenzieren der Bedingung (102) ergibt

$$\frac{dc^{d*}(\theta)}{dr(\theta)} = -\frac{-1}{-H^{-1''}(u - \theta V^*(.))\lambda^*(\theta)^2 + H^{-1'}(u - \theta V^*(.))\frac{\partial \lambda^*(.)}{\partial c^d(\theta)}} < 0. \quad (104)$$

Im Falle eines risikoneutralen Arztes reduziert sich der Ausdruck (104) zu  $\frac{\partial \lambda^*(.)}{\partial c^d(\theta)}$ , der absolute Wert wird also kleiner.

---

<sup>52</sup>Ein häufig beschriebenes Ergebnis lautet, dass der risikoaverse Arzt eine Risikoprämie erhalten muss. Dies ergibt sich, wenn die freien Mittel unsicher sind und eine gewisse Streuung aufweisen, zum Beispiel wenn sie von einem Indikator abhängig gemacht werden, der unsicher mit den Aktionen des Agenten zusammenhängt [Gravelle (2004), Laffont/Martimort (2002)]. Da hier jedoch eine deterministische Umgebung angenommen wird, gibt es auch keine Risikoprämie.

### 6.6.3 Optimale Pauschale $k(\theta)$ und Kostenbeteiligung $r(\theta)$ aus Sicht des Auftraggebers

Grundsätzlich gilt für die optimale Wahl der Parameter  $r(\theta)$  und  $k(\theta)$  dasselbe wie im Falle des risikoneutralen Arztes. Mit der Wahl der optimalen Kostenbeteiligung gibt der Auftraggeber dem Management den Anreiz, das richtige Budget zu wählen, die Pauschale  $k(\theta)$  hat weiterhin keinen Einfluss auf die Budgetierung und wird vom Auftraggeber so gewählt, dass der Nutzenüberschuss des Spitals bei  $g$  liegt. Allerdings muss die Pauschale im Falle des risikoaversen Arztes höher sein.

Weiterhin ist die first best-Lösung zu erreichen, wenn die Bedingung

$$\frac{V_{q_1}^{FB}(\cdot)}{V_{q_2}^{FB}(\cdot)} = \frac{B_{q_1}^{FB}(\cdot)}{B_{q_2}^{FB}(\cdot)}. \quad (105)$$

erfüllt ist. Es besteht allerdings ein wichtiger Unterschied. Während der Auftraggeber im Falle des risikoneutralen Arztes indifferent gegenüber der spitalinternen Aufteilung des Nutzenüberschusses war, ist es für die first best-Lösung bei einem risikoaversen Arzt zusätzlich notwendig, dass der Arzt nicht mehr als seinen Reservationsnutzen  $u$  erhält. Zum Erreichen der first best-Lösung ist der Auftraggeber also darauf angewiesen, dass das Management dem Arzt gerade seinen Reservationsnutzen gewährt.

Die zur first best-Lösung führende Kostenbeteiligung liegt bei einem risikoaversen Arzt etwas oberhalb von 0.5. Da das Management bei gegebener Kostenbeteiligung ein höheres Budget wählt, wenn der Arzt risikoavers ist, steuert der Auftraggeber mit einer höheren Kostenbeteiligung dagegen.

## 6.7 Lösung des Modells bei spitalinterner Verhandlungslösung und Risikoaversion des Arztes

Da im Falle des risikoaversen Arztes die spitalinterne Verteilung des Nutzenüberschusses für den Auftraggeber relevant ist, liegt es auf der Hand, dass die PA-Lösung und die NV-Lösung nicht mehr identisch sein können und die Äquivalenz der beiden internen Mechanismen nicht mehr gegeben ist. Im Vergleich zur PA-Lösung ist der entscheidende Punkt der NV-Lösung, dass der Arzt je nach Verhandlungsmacht ein Nutzenniveau über seinem Reservationsnutzen verhandeln kann. So hängt auch die resultierende Lösung letzten Endes von den internen Machtverhältnissen ab. Für den Auftraggeber bedeutet dies, dass er bei der Bestimmung der optimalen Parameter seines Vergütungssystems  $r(\theta)$  und

$k(\theta)$  Informationen über die internen Machtverhältnisse benötigt.

Die Analyse erfolgt analog zum Falle der Verhandlung mit einem risikoneutralen Arzt. Da die Bestimmung der Mengen  $q_1^*(.)$  und  $q_2^*(.)$  durch den Arzt bei gegebenen Budget unberührt bleiben und die Ergebnisse aus Abschnitt 6.3.2 auch hier gelten, springen wir direkt in die Spielstufe der Verhandlung.

### 6.7.1 Bestimmung der Budgetierung $c^d(\theta)$ sowie der freien Mittel $t^d(\theta)$ durch Verhandlungen zwischen Arzt und Management

Im Rahmen der Verhandlung suchen Management und Arzt eine Lösung  $(t^d(\theta), c^d(\theta))$  für jede Fallschwere. Die Nash-Verhandlungslösung maximiert also

$$\max_{c^d(\theta), t^d(\theta)} \pi(\theta) = G(\theta)^w (U(\theta) - u)^{1-w} \quad (106)$$

für jedes  $\theta \in \Theta$ , wobei  $G(\theta) = k(\theta) - r(\theta)c^d(\theta) - t^d(\theta)$  und  $U(\theta) = H(t^d(\theta)) + \theta V^*(.)$ . Wir gehen davon aus, dass ein eindeutiges internes Maximum existiert. Die Bedingungen erster Ordnung lauten

$$\frac{U(\theta) - u}{G(\theta)} = \frac{1-w}{w} H'(t^{d*}(\theta)) \quad (107)$$

und

$$\frac{\lambda^*(\theta, c^{d*}(\theta))}{H'(t^{d*}(\theta))} = r(\theta). \quad (108)$$

Der gemeinsame Überschuss wird so zwischen Arzt und Management aufgeteilt, dass das Verhältnis der Nutzen der rechten Seite von (107) entspricht. Es zeigt sich, dass die Risikoaversion des Arztes bei der Aufteilung des Nutzenüberschusses zu Lasten des Arztes geht. Formal leitet sich dies wegen  $0 < H'(t^d) \leq 1$  aus Bedingung (107) ab.

Die Bedingung erster Ordnung (108) für die Budgetierung entspricht wegen  $H^{-1'}(.) = \frac{1}{H'(t^d(\theta))}$  der Bedingung erster Ordnung (102) für das optimale Budget bei der PA-Lösung, denn beide Lösungen sind Pareto-effizient. Allerdings unterscheiden sich die resultierenden Budgetierungen, da im Falle des risikoaversen Arztes die Budgetierung nicht mehr unabhängig von den freien Mitteln gewählt werden kann. Bei der PA-Lösung wählt das Management die freien Mittel so, dass der Arzt gerade seinen Reservationsnutzen  $u$  erhält. Bei der Verhandlungslösung erhält der Arzt aber einen positiven Nutzenüberschuss,



so dass sich die beiden Lösungen unterscheiden müssen.

Der Grund für den Unterschied liegt letzten Endes darin, dass bei einem risikoaversen Arzt die Umverteilung vom Gewinn des Managements zum Nutzen des Arztes über die freien Mittel mit einem betragsmässigem Verlust des gemeinsamen Nutzenüberschusses verbunden ist:

$$\frac{dN\ddot{U}(\theta)}{dt^d(\theta)} < 0. \quad (109)$$

Daher erfolgt die Umverteilung zugunsten des Arztes zu einem Teil auch über ein höheres Budget.

**6.7.1.1 Komparative Statik** Die Bedingungen (107) und (108) definieren die verhandelten freien Mittel  $t^{d*}$  und die verhandelte Budgetierung  $c^{d*}$  als Funktionen der Verhandlungsmacht  $w$ , der Kostenbeteiligung  $r(\theta)$  sowie - und dies im Unterschied zum Verhandlungsergebnis im Falle des risikoneutralen Arztes - von der Pauschale  $k(\theta)$ :

$$t^{d*}(\theta) \equiv t^{d*}(\theta, w, r(\theta), k(\theta)) \quad (110)$$

$$c^{d*}(\theta) \equiv c^{d*}(\theta, w, r(\theta), k(\theta)). \quad (111)$$

Die Gleichungen (110) und (111) stellen aus Sicht des Auftraggebers die Anreizverträglichkeitsbedingungen des Spitals dar.

Die drei Parameter  $w, r(\theta), k(\theta)$  beeinflussen das Verhandlungsergebnis folgendermassen:

53

$$\frac{dt^{d*}(\theta, \cdot)}{dw} < 0 \quad \frac{dc^{d*}(\theta, \cdot)}{dw} < 0 \quad (112)$$

$$\frac{dt^{d*}(\theta, \cdot)}{dr(\theta)} \leq \geq 0 \quad \frac{dc^{d*}(\theta, \cdot)}{dr(\theta)} < 0 \quad (113)$$

$$\frac{dt^{d*}(\theta, \cdot)}{dk(\theta)} > 0 \quad \frac{dc^{d*}(\theta, \cdot)}{dk(\theta)} > 0 \quad (114)$$

Wenn der Arzt risikoscheu ist, führt eine grössere Verhandlungsmacht des Arztes  $1-w$  also zu einem höheren Fallbudget. Ebenso führt eine höhere Pauschale  $k(\theta)$  zu einer höheren Budgetierung und entfaltet damit wie der Beteiligungsparameter  $r(\theta)$  eine Anreizwirkung.

---

<sup>53</sup>Die mathematische Herleitung ist sehr umfangreich, sie ist in Appendix A.2 wiedergegeben.

**Ergebnis 6.5** *Werden die freien Mittel und die Budgetierung zwischen Arzt und Management verhandelt, und ist der Arzt risikoscheu, dann steigt das verhandelte Fallbudget mit der Verhandlungsmacht des Arztes.*

**Ergebnis 6.6** *Werden die freien Mittel und die Budgetierung zwischen Arzt und Management verhandelt, und ist der Arzt risikoscheu, dann hat auch die Pauschale  $k(\theta)$  eine Anreizwirkung: Das Fallbudget steigt mit der Pauschale.*

### 6.7.2 Bedeutung der Verhandlungslösung für den Auftraggeber

Der Auftraggeber maximiert seinen Nutzen pro Patient der Fallschwere  $\theta$

$$W(\theta, k(\theta), r(\theta)) = \theta B^*(.) - (1 - r(\theta))c^{d*}(\theta, .) - k(\theta) \quad (115)$$

mit der Wahl der Beteiligung  $r(\theta)$  und der Pauschale  $k(\theta)$  unter Berücksichtigung der Anreizverträglichkeitsbedingungen des Spitals (110) und (111) sowie der Teilnahmebedingung  $N\ddot{U}(\theta) \geq g$ . Wir verzichten auf eine mathematische Analyse der Maximierung, halten jedoch fest, dass wegen der Abhängigkeit des gemeinsamen Überschusses von den freien Mitteln (Ausdruck (109)) die Pauschale  $k(\theta)$  im Falle einer Verhandlungslösung und eines risikoaversen Arztes einen Effekt auf die Budgetierung und die Höhe der freien Mittel hat, so dass eventuell die Teilnahmebedingung des Managements nicht bindet und eine interne Lösung für die Pauschale  $k(\theta)$  existiert. Bei einem risikoneutralen Arzt, dies zur Erinnerung, ist die optimale Pauschale als Randlösung immer so niedrig, dass die Teilnahmebedingung des Managements  $N\ddot{U}(\theta) \geq g$  bindet.

Wenn auch bei einem risikoaversen Arzt die Teilnahmebedingung  $N\ddot{U}(\theta) \geq g$  bindet, ist es aber auch ohne mathematische Analyse möglich, Aussagen darüber zu treffen, wie der Auftraggeber den Einfluss des risikoscheuen Arztes auf das Ergebnis der Verhandlung bewertet. Hierzu vergleichen wir zwei Szenarien: Im ersten Szenario verfüge der Arzt über die Verhandlungsmacht  $1 - w'$ , im zweiten über die Verhandlungsmacht  $1 - w''$ , wobei  $1 - w' > 1 - w''$ . Die Verhandlungsmacht des Arztes ist also in Szenario 1 grösser als in Szenario 2. In Szenario 1 resultiere die aus Sicht des Auftraggebers optimale Allokation  $\{c^{d'}, U'\}$ , mit  $c^{d'} \equiv c^{d'}(\theta, w')$  usw. In Szenario 2 resultiere dagegen  $\{c^{d''}, U''\}$ . Dann muss  $U' > U''$  gelten, da der Arzt bei grösserer Verhandlungsmacht einen grösseren Anteil am gemeinsamen Nutzenüberschuss  $g$  erhält. Wenn die Differenz  $U' - U'' > 0$  nur über die freien Mittel finanziert wird, folgt aus (109) für die Präferenz des Auftraggebers  $\{c^{d'}, U'\} \prec \{c^{d''}, U''\}$ . Da aber die Allokation  $\{c^{d''}, U''\}$  unter Szenario 2 optimal ist, folgt  $\{c^{d'}, U'\} \preceq \{c^{d''}, U''\}$  und damit  $\{c^{d'}, U'\} \prec \{c^{d''}, U''\}$ .

**Ergebnis 6.7** *Wenn im Optimum die Teilnahmebedingung des Managements bindet, dann zieht der Auftraggeber einen Arzt mit geringer Verhandlungsmacht einem Arzt mit grosser Verhandlungsmacht vor.*

Aus Ergebnis 6.7 schliessen wir weiter, dass der Auftraggeber die PA-Lösung immer der Verhandlungslösung vorzieht. Dies folgt, da unter der PA-Lösung der Nutzen des Arztes gerade das Reservationsniveau  $u$  erreicht, also immer niedriger ist als unter der Verhandlungslösung und die Teilnahmebedingung des Managements bindet.

**Ergebnis 6.8** *Wenn der Arzt risikoscheu ist, dann zieht der Auftraggeber die PA-Lösung strikt der Verhandlungslösung vor.*

## 6.8 Fazit Kapitel 6

Die erste Analyse des Modells zeigt, dass es für den Auftraggeber nicht unbedingt eine Rolle spielen muss, ob die Budgetierung und die freien Mittel bei einem spitalinternen PA-Mechanismus vom Management festgelegt werden, oder ob die Budgetierung und die freien Mittel das Ergebnis einer Nash-Verhandlungslösung sind.

Allerdings ist die Äquivalenz der beiden Entscheidungsmechanismen von idealisierenden Annahmen wie der Risikoneutralität des Arztes und des Managements abhängig, die dazu führen, dass der gemeinsame Nutzenüberschuss von Arzt und Management über die freien Mittel zwischen den beiden aufgeteilt werden kann, ohne dass es sich der Nutzenüberschuss dadurch ändert. Ist dies nicht möglich, sind die Lösungen des PA-Mechanismus und des NV-Mechanismus nicht äquivalent, wie wir im Falle eines risikoaversen Arztes gesehen haben. Allerdings sind die Differenzen der PA-Lösung und NV-Lösung umso geringer, je geringer die Ausprägung der Risikoaversion des Arztes zumindest im relevanten Bereich ist. Im Falle eines ‚schwach‘ risikoaversen Arztes, bei dem der marginale Nutzen aus den freien Mittel ‚relativ konstant‘ um den Wert eins liegt, sind die Differenzen der PA-Lösung und der NV-Lösung auch relativ gering, so dass der Aspekt unterschiedlicher spitalinterner Entscheidungsprozesse keine prominente Bedeutung hat.

Ähnliches gilt für die Irrelevanz der internen asymmetrischen Verteilung der Information bezüglich der Mengen  $q_1$  und  $q_2$ , die aus der Annahme folgt, dass das Management keine eigenen (direkten) Präferenzen gegenüber den Mengen hat. Wenn das Management über eigene Präferenzen verfügt, die aber in dem Sinne relativ schwach ausgeprägt sind, dass der marginale Nutzen aus den Mengen nahe bei null liegt, spielt die interne Informationsasymmetrie zwischen Arzt und Management zwar auch eine Rolle, aber keine allzu grosse.

## 7 Massnahmen des Managements zur Verbesserung der Kosteneffizienz

### 7.1 Hintergrund

Die Effizienz der Leistungserbringung der Spitäler rückt mit der zunehmenden Diskussion über die steigenden Gesundheitskosten immer weiter in den Mittelpunkt der öffentlichen Wahrnehmung.<sup>54</sup> Die Kosteneffizienz<sup>55</sup> eines Spitals misst sich in dem Verhältnis der Kosten der Leistungserbringung zu dem aus der Leistungserbringung resultierenden Output. Ein Spital behandelt also einen Patienten umso kosteneffizienter, je weniger Kosten es für die Behandlung bei gleichem Behandlungsergebnis verursacht.

Ineffizienzen bei der Behandlung entstehen im Spital durch nicht-optimalen Einsatz von Ressourcen oder Verschwendung [McKay (2008)]. Sie sind auf der einen Seite Ausdruck von Fehlentscheidungen bei der Wahl der Mittel zur Behandlung eines Patienten und daher zu einem grossen Teil vom Arzt zu verantworten. Der Beitrag des Arztes zur *Vermeidung* von Kostenineffizienzen liegt also vor allem darin, das Bündel an Massnahmen zur Behandlung eines Patienten einzusetzen, welches unter der Bedingung der „medizinischen Angemessenheit“ die, bei gegebenen Preisen, günstigste „Kosten-Nutzen-Relation“ aufweist [Frese et al. (2004)]. Im PA-Modell mit doppeltem Unterstellungsverhältnis wird dieser Aspekt (aus Sicht des Auftraggebers) durch das Verhältnis der Bewertung der Behandlung  $\theta B(q_1, q_2)$  zu den Behandlungskosten  $c = p_1 q_1 + p_2 q_2$  bereits berücksichtigt.

Auf der anderen Seite sind Ineffizienzen vor allem aber auch Ausdruck von Defiziten bei der Optimierung, Koordination und Überwachung der spitalinternen Prozesse und daher auch dem Verantwortungsbereich des Managements zuzuordnen. Der Beitrag des Managements zur Vermeidung von Ineffizienzen besteht darin, einen gegebenen mengenmässigen Ressourcenverbrauch zu niedrigeren Kosten zu ermöglichen. Im Modell bedeutet dies also bei gegebenen Mengen  $q_1$  und  $q_2$  die Kosten  $c$  zu senken.

Das Potential von geeigneten Managementmassnahmen zur Verbesserung der Kosteneffi-

---

<sup>54</sup>Selbst der Gesetzgeber in der Schweiz fühlte sich veranlasst, die effiziente Leistungserbringung im Zuge der Revision des KVG (Krankenversicherungsgesetz) 2007 im Gesetz zu verankern, wenngleich in recht schwammiger Form (KVG Art. 49).

<sup>55</sup>Der Begriff der Kosteneffizienz ist zu unterscheiden vom Begriff der technischen Effizienz. Bei der technischen Effizienz wird ein ggf. mehrdimensionaler Vektor von Produktionsfaktoren (in dem hier diskutierten Zusammenhang z.B. Anzahl Ärzte, Summe der zur Pflege aufgebrauchten Minuten und Milliliter einer Desinfektionslösung) den daraus produzierten Gütern gegenübergestellt. Bei der Kosteneffizienz dagegen wird die Summe der zu Marktpreisen bewerteten Produktionsfaktoren relevant [vgl. Steinmann (2003), Seite 1].

zienz wird in der Literatur aus der Spitalpraxis immer wieder erwähnt [z.B. Goldschmidt (2009), Radl (2009) oder Palm (2009)]. Als ein wichtiger Bereich in der Praxis ist die interdisziplinäre Planung des gesamten Ablaufs der Behandlung des Patienten von der Aufnahme bis zur Abrechnung zu nennen. Durch solche unter dem Schlagwort „Optimierung der Patientenpfade“ im Spitalmanagement bekannte ganzheitliche Abstimmung und Koordination der internen Abläufe konnten in der Praxis nachweislich erhebliche Effizienzgewinne erzielt werden [Palm (2009)]. Ein in der Realität häufig anzutreffendes Beispiel ist hier das OP-Team, das am OP-Tisch wartet, während der Operateur noch bei einem anderen Patienten tätig ist. Solcherlei Leerläufe und Ineffizienzen sind das Ergebnis unzureichender Koordination des OP-Betriebes. Durch entsprechende Management-Massnahmen, wie zum Beispiel ein EDV-gestütztes OP-Management, sind in der Praxis bereits wesentlich reibungsfreiere Abläufe realisiert worden [Radl (2009)].

Ein weiterer Bereich für Massnahmen des Managements zur Verbesserung der Kosteneffizienz, der häufig erwähnt wird, ist die Koordination des Einkaufs von medizinischem Material und Medikamenten. Durch ein zentrales Einkaufsmanagement oder auch die Bildung von Einkaufsgemeinschaften mit anderen Spitälern können niedrigere Einkaufspreise und Rabatte bei den Lieferanten erzielt werden als dies bei einer dezentralen Beschaffung durch die einzelnen klinischen Bereiche des Spitals der Fall wäre.

Für das Spitalmanagement sind derartige Massnahmen allerdings mit Kosten und Arbeitsbelastung verbunden. Die Kosten bei der Umsetzung von Massnahmen zur Effizienzsteigerung bestehen beispielsweise in den Aufwendungen für externe Beratungen oder dem Einkauf und der Pflege von EDV-Systemen. Eine höhere Arbeitsbelastung hingegen resultiert zum Beispiel aus der durch die Massnahmen bedingten Mehrarbeit oder der Überwindung von spitalinternen Widerständen bei der Umsetzung, die in der Regel entstehen, sobald etwas Neues eingeführt wird.

Auf der anderen Seite sind die spitalinternen Massnahmen für einen Auftraggeber und Finanzier der Spitalleistungen vertraglich kaum festzulegen. Zum einen müsste er dazu wissen, welche Massnahmen zu ergreifen sind. Diese detaillierten spitalinternen Informationen werden ihm aber als Aussenstehenden nicht zur Verfügung stehen. Zum anderen wird der Erfolg der Massnahmen von der Intensität der Massnahmen und dem Arbeitseinsatz des Managements bestimmt. Auch in dieser Hinsicht ist eine vertragliche Regelung schwer vorstellbar.

Aus diesem Sachverhalt lässt sich folgern:

**Folgerung 7.1** *Da die Durchführung der Massnahmen zur Verbesserung der Kosteneff-*

*fizienz zum einen für das Management mit Kosten und Arbeitsbelastung verbunden, zum anderen vom Auftraggeber vertraglich kaum festzulegen sind, bedarf es geeigneter Anreize für das Management, die Massnahmen in geeigneter Intensität durchzuführen.*

## **7.2 Asymmetrische Information über die Anstrengungen zur Verbesserung der Kosteneffizienz in der Theorie der Spitalvergütung**

In der Theorie stellt sich aus Sicht des Auftraggebers das optimale Niveau der Anstrengungen des Spitals zur Steigerung der Kosteneffizienz dann ein, wenn die marginale Ersparnis bei den Behandlungskosten gerade den marginalen Kosten entspricht, die dem Spital durch die Anstrengung entstehen. Ein allgemein gültiges Ergebnis lautet hier, dass das Spital das first best-Niveau der Anstrengungen zur Verbesserung der Kosteneffizienz wählt, wenn der Auftraggeber dem Spital die Kostenverantwortung zu 100% überträgt [vgl. Shleifer (1985) oder Breyer/Zweifel/Kifmann (2005)]. Allerdings wirken sich die mit der Kostenverantwortung verbundenen Anreize zur Kostensenkung auch negativ auf die Bereitstellung der Behandlungsqualität aus. Hier entsteht für den Auftraggeber grundsätzlich ein Trade-off zwischen der Kosteneffizienz und der Qualität der Behandlung, wenn ihm zur Beeinflussung der Leistungserbringung des Spitals lediglich die beobachteten Behandlungskosten als Instrument zur Verfügung stehen. Dem Auftraggeber ist es dann nicht möglich zu beurteilen, ob eine Kostensenkung die Folge von verbesserter Kosteneffizienz ist, oder ob das Spital lediglich an der Qualität Einsparungen vorgenommen hat.

Dieser Trade-off zwischen Kosteneffizienz und Qualität der Behandlung ist ein zentraler Punkt in der theoretischen Literatur über die Vergütung von Spitälern. Zu erwähnen ist zum Beispiel der in Kapitel 4 diskutierte Artikel von Chalkley und Malcomson [Chalkley/Malcomson (1998a)]. Dort zeigen die Autoren, dass es für den Auftraggeber optimal sein kann, dem Spital nur teilweise die Kostenverantwortung zu übertragen und einen Kompromiss zwischen optimaler Kosteneffizienz und einer Qualitätsbereitstellung einzugehen, die über einem vertretbaren Mindestmass liegt. Allerdings bedarf es hierzu eines Eigeninteresses des Spitals an der Behandlungsqualität. Verfolgt das Spital dagegen lediglich das Ziel der Maximierung des finanziellen Gewinns, wird es nur bei einer vollständigen retrospektiven Kostenerstattung (eine Kostenverantwortung von null also) bereit sein, ein Qualitätsniveau bereitzustellen, dass über einem (verifizierbaren) Mindestmass liegt.

Wenn dem Auftraggeber allerdings weitere Informationen über die Leistungserstellung des Spitals zugänglich sind, kann die Berücksichtigung dieser Zusatzinformationen in der Vergütung des Spitals eine Verbesserung des Ergebnisses ermöglichen bzw. der Trade-off zwischen Kosteneffizienz und Behandlungsqualität gemildert und im optimalen Fall sogar

komplett eliminiert werden. So nimmt zum Beispiel Ma (1995) in seinem Modell an, dass die Nachfrage der Patienten nach der Behandlung im Spital mit der Qualität der Behandlung steigt. Ein Spital, das nur an der Maximierung des finanziellen Gewinnes interessiert ist, wird genau das Niveau an Behandlungsqualität wählen, bei dem die marginalen Kosten aus der zusätzliche Einheit an Qualität und der Behandlung des zusätzlichen Patienten, der die Behandlung nachfragt, dem marginalen Ertrag aus dem zusätzlichen Patienten entspricht. Wenn der Auftraggeber die Nachfrage kennt, kann er sich diese Information zunutze machen, indem er als Vergütung eine Pauschale pro Fall definiert, die den marginalen Kosten bei first best-Behandlungsqualität entspricht. Da das Spital die volle Kostenverantwortung trägt, wählt es ausserdem das first best-Niveau der Anstrengung zur Verbesserung der Kosteneffizienz. Der Auftraggeber kann in diesem Fall den Trade-off zwischen Kosteneffizienz und Behandlungsqualität vollständig eliminieren und erreicht die first best-Allokation. Ma nimmt allerdings an, dass die resultierende Anzahl der Patienten auch die vom Auftraggeber gewünschte ist und zudem das Spital die notwendigen Kapazitäten hat.

In einem zweiten Artikel von Chalkley und Malcomson wird das Modell von Ma unter anderem in der Hinsicht erweitert, dass die Nachfrage bei der first best-Qualität nicht unbedingt der optimalen Anzahl der behandelten Patienten entspricht [Chalkley/Malcomson (1998b)]. In diesem Zusammenhang lassen Chalkley und Malcomson zwei weitere Szenarien zu: Zum einem betrachten sie den Fall, dass die aus Sicht des Auftraggebers optimale Anzahl der zu behandelnden Patienten grösser ist als die Kapazität des Spitals. Zum andern erörtern sie den Umstand, wenn die first best-Anzahl an zu behandelnden Patienten kleiner ist als die Anzahl von Patienten, die bei der first best-Qualität Behandlung nachfragen. Wenn die Qualität eindimensional ist, zeigen Chalkley und Malcomson, dass es auch in diesen Fällen für den Auftraggeber möglich ist, die first best-Qualität und die first best-Anstrengung zu erreichen. Hierfür zahlt der Auftraggeber dem Spital zusätzlich zur Pauschale für einen behandelten Patienten eine weitere Pauschale für jeden Patienten, der die Behandlung im Spital nachfragt. Während der Auftraggeber mit der optimalen Pauschale für jeden tatsächlich behandelten Patient dem Spital Anreize gibt, die richtige Anzahl an Patienten zur Behandlung aufzunehmen, erreicht der Auftraggeber mit einer Pauschale für jeden Patienten, der Behandlung nachfragt, eine optimale Bereitstellung der Qualität. Da dieses System keine cost-sharing Komponente aufweist, dem Spital also die Kostenverantwortung zu 100% überträgt, wählt das Spital auch das first best-Niveau der Anstrengungen zur Verbesserung der Kosteneffizienz, der Auftraggeber muss also keinen Kompromiss zwischen Qualität und Kosteneffizienz finden.

### 7.3 Massnahmen des Managements zur Verbesserung der Kosteneffizienz im Prinzipal-Agent-Modell mit doppeltem Unterstellungsverhältnis

In unserem PA-Modell mit doppeltem Unterstellungsverhältnis sind die Kosten der Behandlung mit  $c = p_1 q_1 + q_2 p_2$  definiert. Wenn sich die Massnahmen des Managements zur Verbesserung der Kosteneffizienz dazu führen, dass dadurch die Behandlungskosten bei gegebenen Mengen  $q_1$  und  $q_2$  sinken, müssen sich die Massnahmen im Modell auf die Preise  $p_1$  und  $p_2$  auswirken.

Wie in Abschnitt 5.4 erwähnt, werden die Preise  $p_1$  und  $p_2$  als ‚Bruttopreise‘ interpretiert, die neben den Herstellerpreisen auch spitalinterne Zuschläge, wie zum Beispiel den Verwaltungsaufwand, enthalten. Die Massnahmen des Managements können also einerseits bei der Beeinflussung des Marktpreises ansetzen, zum Beispiel durch Erwirken und Ausnutzen von Lieferantenrabatten durch geschicktes Einkaufsmanagement, andererseits aber auch die spitalinternen Zuschläge senken, indem zum Beispiel die Prozesse optimiert werden wie oben erläutert.

Grundsätzlich könnte man, wie im Falle der Behandlungsquantitäten, auch bei der Kostensenkung von einem *mehrdimensionalen* Bündel an Massnahmen ausgehen, der *beide* Preise  $p_1$  und  $p_2$  beeinflusst. Dadurch würde aber die Analyse deutlich komplexer, für die hier diskutierte Problematik der Wirkungsweise der internen Mechanismen wären dadurch jedoch keine entscheidenden Einsichten gewonnen. So beschränken wir uns hier auf den Fall, dass dem Management lediglich ein eindimensionaler Parameter  $e \geq 0$  zur Verfügung steht, der auch nur den Preis für die Leistung 1,  $p_1$ , beeinflusst. Der Parameter  $e$  ist dabei als ein Indikator für die Anstrengungen zu verstehen, die das Management zum Erreichen eines niedrigeren Preises  $p_1$  unternimmt.

Wir nehmen an, dass zwischen der Anstrengung  $e$  und dem Preis  $p_1$  ein negativ-konvexer Zusammenhang besteht:

$$p_1 = p_1(e), \text{ mit } p_1'(e) < 0, \text{ und } p_1''(e) > 0. \quad (116)$$

Höhere Anstrengungen resultieren in einem niedrigeren Preis für eine Einheit der Leistung  $q_1$ . Allerdings müssen die zur marginalen Senkung des Preises  $p_1$  notwendigen marginalen Anstrengungen umso höher sein, je niedriger der Preis schon ist. Erbringt das Management dagegen keine Anstrengungen, resultiert ein maximaler Preis  $p_1(0) = p_1^{max} > 0$ . Auf der anderen Seite nehmen wir an, dass es dem Management nicht möglich ist, den Preis auf null zu senken, indem wir



$$\lim_{e \rightarrow \infty} p_1(e) \rightarrow 0 \quad (117)$$

annehmen.

Die Anstrengungen, die das Management zur Senkung der Preise unternimmt, sind mit persönlichen Kosten oder Arbeitsleid für das Management verbunden. Die Kosten und das Arbeitsleid spiegeln die mit den Massnahmen verbundene Belastung durch Mehrarbeit, Überwinden von internen Widerständen gegen Veränderungen etc. wider. Wir bezeichnen die Kosten des Arbeitsleides mit  $\varphi(e)$  und nehmen an,  $\varphi(e)$  ist konvex und monoton ansteigend<sup>56</sup>, mit  $\varphi(0) = 0$  und  $\lim_{e \rightarrow \infty} \varphi(e) \rightarrow \infty$ .

Um die Darstellung übersichtlicher zu halten, stellen wir im Folgenden die Kosten des Arbeitsleides als direkte Funktion des Preises dar:<sup>57</sup>

$$\varphi(e(p_1)) \equiv \varphi(p_1). \quad (118)$$

In Anhang A.3 wird gezeigt, dass  $\varphi(p_1)$  eine konvex fallende Funktion ist, für die gilt:

$$\varphi'(p_1) < 0, \quad \varphi''(p_1) > 0 \text{ und } \varphi(p_1^{max}) = 0, \quad \lim_{p \rightarrow 0} \varphi(p) \rightarrow \infty. \quad (119)$$

Wir nehmen an, das Management leiste die Anstrengungen *bevor* die Patienten zur Behandlung ins Spital kommen, damit der Arzt die Behandlung, wie im vorigen Kapitel, bei *gegebenen* Preisen durchführen kann. Da der Preis  $p_1$  für die Behandlung aller  $n$  Patienten gilt, wählt das Management die Anstrengungen  $e$  so, dass sein erwarteter Gewinn aus der Behandlung aller  $n$  Patienten maximiert wird:

$$nEG = n \int_{\underline{\theta}}^{\bar{\theta}} [k(\theta) - r(\theta)c^d(\theta) - t^d(\theta)]f(\theta)d(\theta) - \varphi(p_1). \quad (120)$$

Die Zielfunktion lässt sich mit  $\bar{\varphi}(p_1) \equiv \varphi(p_1)/n$  auch pro Patient formulieren

$$EG = \int_{\underline{\theta}}^{\bar{\theta}} [k(\theta) - r(\theta)c^d(\theta) - t^d(\theta)]f(\theta)d(\theta) - \bar{\varphi}(p_1). \quad (121)$$

<sup>56</sup>Dies ist eine gängige Annahme in der Prinzipal-Agent-Literatur (z.B. Laffont/Tirole 1993).

<sup>57</sup>Im Folgenden wird des Öfteren davon die Rede sein, dass das Management den Preis  $p_1$  wählt. Dies geschieht des einfacheren Ausdrucks wegen. Natürlich ist aber damit gemeint, dass das Management die Intensität der Anstrengungen bestimmt, woraus dann der effektive Preis resultiert.

## 7.4 Asymmetrische Information hinsichtlich der Anstrengungen das Managements und des Preises $p_1$ im Modell mit doppeltem Unterstellungsverhältnis

### 7.4.1 Annahmen an die Informationsverteilung

Wir nehmen an, dass mit den Anstrengungen bzw. mit dem aus den Anstrengungen resultierenden effektiven Preis  $p_1$  eine weitere Grösse neben den Mengen  $q_1$  und  $q_2$  besteht, deren Ausprägung der Auftraggeber nicht beobachten und vertraglich festlegen kann.

Tabelle 2 gibt einen Überblick über die Verteilung der Information in diesem Kapitel.

### 7.4.2 First best-Allokation

Wie im vorigen Kapitel bestimmen wir zunächst die first best-Lösung, die der Auftraggeber bei vollständiger Information wählen würde.

Da der Preis  $p_1$  für die Behandlung *aller*  $n$  Patienten gilt, der Preis  $p_1$  aber festgelegt wird, bevor die konkreten Fallschweren bekannt werden, lautet das Maximierungsprogramm des Auftraggebers für die first best-Lösung:

$$\max_{p_1, q_1(\theta), q_2(\theta), N\ddot{U}(\theta)} nEW = n \int_{\underline{\theta}}^{\bar{\theta}} [(\theta B(.) + \theta V(.) - p_1 q_1(\theta) - p_2 q_2(\theta)) - N\ddot{U}(\theta)] f\theta d\theta - \varphi(p_1), \quad (122)$$

oder äquivalent pro Patient

$$\max_{p_1, q_1(\theta), q_2(\theta), N\ddot{U}(\theta)} EW = \int_{\underline{\theta}}^{\bar{\theta}} [(\theta B(.) + \theta V(.) - p_1 q_1(\theta) - p_2 q_2(\theta)) - N\ddot{U}(\theta)] f\theta d\theta - \bar{\varphi}(p_1). \quad (123)$$

<i>Grösse beobachtbar ja/nein</i>	Auftraggeber	Management	Arzt
Mengen $q_1, q_2$	nein	nein	ja
Anstrengung $e$ bzw. Preis $p_1$	nein	ja	ja
Preis $p_2$	ja	ja	ja
Kosten $c$	ja	ja	ja
Fallschwere $\theta$	ja	ja	ja

Tabelle 2: Informationsverteilung des Modells in Kapitel 7.

Da der gemeinsame Nutzenüberschuss negativ in die Zielfunktion des Auftraggebers eingeht, ist es für ihn optimal, wenn der gemeinsame Nutzenüberschuss von Arzt und Management minimal ist

$$NU(\theta) = g. \quad (124)$$

Die spitalinterne Verteilung des Nutzenüberschusses zwischen Management und Arzt spielt dagegen keine Rolle.

Die Bedingungen erster Ordnung für die optimalen Pfade von  $q_1(\theta)$  und  $q_2(\theta)$  erhalten wir durch punktweise Maximierung von (123).<sup>58</sup> Sie entsprechen denjenigen aus Abschnitt 6.2.2:

$$\theta B_{q_1}^{FB}(\cdot) = p_1^{FB} - \theta V_{q_1}^{FB}(\cdot) \quad (125)$$

$$\theta B_{q_2}^{FB}(\cdot) = p_2 - \theta V_{q_2}^{FB}(\cdot). \quad (126)$$

Hinsichtlich des Preises  $p_1$  ergibt sich als Bedingung erster Ordnung

$$Eq_1^{FB}(\theta) = -\bar{\varphi}'(p_1^{FB}) \quad (127)$$

Die durch eine marginale Senkung des Preises erwartete Kostensenkung muss den marginalen Kosten der Anstrengung entsprechen. Wir gehen davon aus, dass die Bedingungen zweiter Ordnung erfüllt sind.

## 7.5 Analyse des Modells bei spitalinterner Prinzipal-Agent-Lösung

### 7.5.1 Zeitliche Abfolge

Wir nehmen an, das Management wähle die Preise nach dem Vertragsangebot des Auftraggebers, aber vor der Behandlung der Patienten durch den Arzt. Es ergibt sich also folgender zeitlicher Ablauf:

1. In der ersten Stufe des Spieles bestimmt der Auftraggeber die Parameter  $k(\theta)$  und  $r(\theta)$  für jede Fallschwere  $\theta \in \Theta$ .

---

<sup>58</sup>Obwohl mit  $q(\theta)$  ein optimaler Pfad gewählt wird, ist dies kein wirkliches Problem dynamischer Maximierung. Die Wahl von  $q_1$  bzw.  $q_2$  zu einer bestimmten Fallschwere  $\theta$  beeinflusst nicht die Wahl bei einer anderen Fallschwere. Von daher entspricht die Maximierung von (123) einer Maximierung des Integranden für jedes  $\theta$ .

2. In der zweiten Stufe entscheidet das Management über die Annahme oder die Ablehnung des Vertragsangebotes des Auftraggebers. Wenn es annimmt, leistet das Management die Anstrengungen zur Senkung des Preises  $p_1$  und bestimmt für jede Fallschwere  $\theta \in \Theta$  die Fallbudgetierung  $c^d(\theta)$  und die freien Mittel  $t^d(\theta)$  für den Arzt.
3. In der dritten Stufe kommen die  $n$  Patienten zur Behandlung in das Spital. Der Arzt stellt die jeweilige Fallschwere  $\theta$  der einzelnen Patienten fest und entscheidet, ob er den Patienten zur Behandlung aufnimmt oder ob er ihm die Aufnahme verweigert. Wenn er sich für die Behandlung des Patienten entscheidet, berichtet der Arzt die Fallschwere wahrheitsgemäss an das Management. Das Management entscheidet, ob es die Behandlung zulässt. Wenn das Management die Behandlung zulässt, bestimmt der Arzt die Mengen  $q_1, q_2$  für die Behandlung des Patienten so, dass das vom Management vorgegebene Budget  $c^d(\theta)$  eingehalten wird. Das Management berichtet die Fallschwere wahrheitsgemäss an den Auftraggeber. Die Behandlungskosten werden von allen drei Akteuren beobachtet. Die Zahlungen finden statt.

Wie üblich lösen wir das Modell mittels Rückwärtsinduktion und beginnen bei der letzten Stufe.

### 7.5.2 Bestimmung der Mengen $q_1$ und $q_2$ durch den Arzt

Der Arzt bestimmt die Mengen bei gegebenen Preisen und gegebenem Budget, so dass die Ergebnisse aus Abschnitt 6.3.2 auch hier gültig sind. Es resultieren also die Bedingungen erster Ordnung (48) - (50), es wird aber betont, dass die optimalen Mengen  $q_1^*$  und  $q_2^*$  sowie der Lagrange-Multiplikator  $\lambda^*(\theta, \cdot)$  auch vom Preis  $p_1$  abhängen. Im Folgenden gilt daher die vereinfachende Notation

$$q_1^*(\cdot) \equiv q_1^*(c^d(\theta), p_1), \quad q_2^*(\cdot) \equiv q_2^*(c^d(\theta), p_1), \quad \lambda^*(\theta, \cdot) \equiv \lambda^*(\theta, c^d(\theta), p_1, \cdot). \quad (128)$$

Allerdings sind für die weitere Analyse Annahmen bzw. Aussagen darüber notwendig, wie der Arzt auf eine Änderung des Preises  $p_1$  reagiert und mit welchem Betrag der Arzt bei einer Änderung des Preises  $p_1$  kompensiert werden muss, wenn durch die Änderung der Gesamtnutzen des Arztes konstant bleiben soll.

**7.5.2.1 Anpassung der Mengen an eine Änderungen des Preises  $p_1$**  Bei der Bestimmung des Preises  $p_1$  muss das Management die Reaktion des Arztes bei der Wahl der Mengen beachten. Es sei angenommen, der Arzt senkt bei der Behandlung eines

Patienten der Fallschwere  $\theta$  bei gegebenem Budget  $c^d(\theta)$  die Menge  $q_1^*(.)$ , wenn der Preis  $p_1$  steigt:<sup>59</sup>

$$\frac{\partial q_1^*(.)}{\partial p_1} < 0 \quad (129)$$

Den Einfluss des Preises auf die Menge  $q_2^*(.)$  lassen wir unbestimmt:

$$\frac{\partial q_2^*(.)}{\partial p_1} < > 0. \quad (130)$$

### 7.5.3 Kompensation des Arztes für eine Änderung des Preises $p_1$

Eine Erhöhung des Preises  $p_1$  wirkt sich bei gegebener Budgetierung  $c^d(\theta)$  negativ auf den erreichbaren Nutzen des Arztes aus der Behandlung eines Patienten aus. Mit Hilfe des Envelope - Theorems angewandt auf die Optimalwertfunktion  $\Phi(.)$  des Arztes

$$\Phi(\theta, c^d(\theta), p_1) = \theta V^*(.) + \lambda^*(\theta, .)(c^d(\theta) - p_1 q_1^*(.) - p_2 q_2^*(\theta, .)) \quad (131)$$

ergibt sich der Nutzenverlust durch eine Erhöhung des Preises  $p_1$ :

$$\frac{d\Phi(\theta, c^d(\theta), p_1)}{dp_1} = -\lambda^*(\theta, c^d(\theta), p_1) q_1^*(.) < 0. \quad (132)$$

Wenn der Nutzen des Arztes keine Änderung erfahren soll, muss der (risikoneutrale) Arzt für die Erhöhung des Preises  $p_1$  mit einer Erhöhung der freien Mittel um den Betrag  $\lambda^*(\theta, .) q_1^*(.) > 0$  kompensiert werden.

### 7.5.4 Bestimmung des Preises $p_1$ , der Budgetierung $c^d(\theta)$ sowie der freien Mittel $t^d(\theta)$ durch das Management.

Wie bereits erwähnt, impliziert der zeitliche Ablauf, dass das Management das Budget  $c^d(\theta)$  und den Preis  $p_1$  festlegen muss, bevor die Behandlung der Patienten und damit der Verbrauch der Mengen durch den Arzt stattfindet. Das Management maximiert also den Erwartungsgewinn pro Patient (121) mit der Bestimmung der Budgetierung, der freien Mittel und des Preises  $p_1$ . Mit dem Nutzen des Arztes  $U(\theta) = t^d(\theta) + V^*(.)$  lässt sich das Maximierungsprogramm des Managements auch äquivalent darstellen:

---

<sup>59</sup>Im Sinne der Konsumententheorie schliessen wir also den extremen Fall aus, dass Leistung 1 für den Arzt die Bedeutung eines Giffen-Gutes hat und der Arzt auf eine Erhöhung des Preises mit einer Erhöhung der Menge  $q_1^*(.)$  reagiert [vgl. Breyer (2004)].

$$\max_{U(\theta), c^d(\theta), p_1} E[G(\theta)] = \int_{\underline{\theta}}^{\bar{\theta}} [k(\theta) - r(\theta)c^d(\theta) + \theta V^*(.) - U(\theta)] f(\theta) d(\theta) - \bar{\varphi}(p_1) \quad (133)$$

wobei wieder  $\bar{\varphi}(p_1) \equiv \varphi(p_1)/n$ . Da der Nutzen des Arztes negativ in die Zielfunktion eingeht, gewährt ihm das Management gerade den Reservationsnutzen  $u$ . Hinsichtlich des optimalen Preises  $p_1$  und des „Pfades“ der Budgetierung  $c^d(\theta)$  entlang der Fallschwere, lösen wir das Problem (133), in dem wir den Maximierungsprozess gedanklich in zwei Stufen zerlegen. In der ersten Stufe maximiert das Management die Zielfunktion mit der Bestimmung der Budgetierung  $c^d(\theta)$  für jede Fallschwere bei gegebenem Preis  $p_1$ . Punktweise Maximierung liefert hier

$$\lambda^*(\theta, c^d(\theta), p_1) - r(\theta) = 0. \quad (134)$$

Gleichung (134) definiert die Budgetierung für einen Patienten mit der Fallschwere  $\theta$  als eine Funktion des Preises  $p_1$  und der Kostenbeteiligung  $r(\theta)$ :

$$c^{*d}(\theta) \equiv c^d(\theta, r(\theta), p_1). \quad (135)$$

In der zweiten Stufe maximiert das Management die Zielfunktion (133) mit der Wahl des Preises  $p_1$  bei gegebener Budgetfunktion  $c^{*d}(\theta, r(\theta), p_1)$ . Die Bedingung erster Ordnung lautet

$$-E[\lambda^*(.)q_1^*(.) + (r(\theta) - \lambda^*(.)) \frac{\partial c^{*d}(.)}{\partial p_1}] - \bar{\varphi}'(p_1^*) = 0. \quad (136)$$

Mit (134) vereinfacht sich (136) zu

$$-E[r(\theta)q_1^*(p_1^*, .)] - \bar{\varphi}'(p_1^*) = 0. \quad (137)$$

Das Management bestimmt den Preis  $p_1^*$  so, dass die erwartete Kosteneinsparung dem marginalen Nutzenverlust durch die Anstrengungen entspricht.<sup>60</sup>

### 7.5.5 Maximierungsproblem des Auftraggebers: Trade-off zwischen Kosteneffizienz und Fallbudgetierung

Das Maximierungsproblem des Auftraggebers lautet in der äquivalenten Schreibweise mit dem Nutzenüberschuss als zu bestimmende Grösse

---

<sup>60</sup>Wir nehmen an, dass die Bedingungen zweiter Ordnung des Maximierungsproblems erfüllt sind.

$$\max_{r(\theta), N\ddot{U}(\theta)} E[W(\theta)] = \int_{\underline{\theta}}^{\bar{\theta}} [(\theta B(\cdot) + \theta V(\cdot) - c^d(\theta) - N\ddot{U}(\theta))] f\theta d\theta - \bar{\varphi}(p_1), \quad (138)$$

unter Berücksichtigung der Anreizverträglichkeitsbedingungen für das Budgetverhalten (134) des Managements und der gewählten Anstrengungen zum Erreichen des Preises  $p_1$  (137) sowie der Anreizverträglichkeitsbedingungen des Arztes (128).

Die Lösung des Problems ist komplizierter als diejenige in Abschnitt 6.3.6 bei exogen gegebenem Preis  $p_1$ . Die Beteiligung beeinflusst nun nicht nur die Budgetierung, sondern auch den Preis  $p_1$ . Beide haben wiederum jeweils einen Effekt auf die Mengen  $q_1$  und  $q_2$ . Hinzu kommt, dass für die Budgetierung eines Falles der Fallschwere  $\theta$ ,  $c^d(\theta, \cdot)$  wegen (137) nun nicht mehr nur die Beteiligung für diese Fallschwere  $r(\theta)$  entscheidend ist, sondern die Beteiligungen für alle andere Fallschweren. Da jedoch aus der mathematischen Analyse ohne weitere Annahmen keine eindeutigen Ergebnisse folgen, wird hier davon abgesehen und die grundsätzliche Problematik verbal diskutiert.

Wenn das Management den Preis  $p_1$  durch entsprechende Anstrengungen beeinflussen kann, der Preis aber gleichzeitig für den Auftraggeber nicht zu beobachten ist, besteht für den Auftraggeber lediglich die Möglichkeit, mit der Beteiligung des Managements an den Kosten dessen Anstrengungen und damit den Preis zu beeinflussen. Von der Pauschale gehen keine Anreize aus. Dieser Versuch steht aber in Konflikt mit dem Versuch, über die Kostenbeteiligung dem Management Anreize zur optimalen Budgetierung zu geben. Es entsteht für den Auftraggeber also ein grundsätzlicher Trade-off zwischen optimaler Kosteneffizienz und optimaler Budgetierung bzw. zwischen optimaler Kosteneffizienz und optimalem Behandlungsnutzen.

Wenn zum Beispiel  $\theta B(q_1, q_2) \equiv \theta V(q_1, q_2)$  gilt und Arzt und Management die Mengen identisch bewerten, lässt sich dies besonders deutlich illustrieren. Aus Abschnitt 6.3.6 wissen wir, dass bei gegebenem first best-Preis der Auftraggeber mit einer Kostenbeteiligung von  $r(\theta) = 0.5$  für alle Fallschweren die first best-Mengen erreicht. Bei gegebenen first best-Mengen folgt aus einem Vergleich der Bedingungen (137) und (127), dass das Management nur dann den first best-Preis wählt, wenn  $E[r(\theta)q_1^*(\cdot)] = E[q_1^*(\cdot)]$  erfüllt ist, was bei  $r(\theta) = 1$  für alle Fallschweren gilt. Eine Kostenbeteiligung  $r(\theta) = 0.5 \forall \theta$ , die für die first best-Mengen notwendig ist, beinhaltet daher keine ausreichenden Anreize für das Management, den first best-Preis zu wählen. Auf der anderen Seite führt eine Beteiligung  $r(\theta) > 0.5 \forall \theta$  zu einer zu niedrigen Budgetierung, um die first best-Mengen zu erreichen. Der Trade-off, den der Auftraggeber hier lösen muss, besteht in möglichst hohen Anreizen zur kosteneffizienten Behandlung einerseits und ausreichender Allokation der Mengen

andererseits. Die first best-Lösung ist in keinem Fall mehr möglich.

**Ergebnis 7.1** *Wenn die Mengen  $q_1$  und  $q_2$  und der Preise  $p_1$  vom Auftraggeber vertraglich nicht festzulegen sind, ist bei der PA-Lösung die first best-Allokation nicht zu erreichen. Der Auftraggeber muss einen Trade-off zwischen Kosteneffizienz und Budgetierung eingehen.*

## 7.6 Analyse des Modells bei spitalinterner Verhandlungslösung

Zunächst stellt sich die Frage, ob der endogene Preis  $p_1$  innerhalb der Verhandlung zwischen Arzt und Management festgelegt wird, oder ob das Management den Preis ohne direkten Einfluss des Arztes ausserhalb der Verhandlungen bestimmt. Im Hinblick auf die Praxis gehen wir von letzteren aus. Am Universitätsspital Zürich zum Beispiel dreht es sich bei den Budgetierungsgesprächen und Verhandlungen mit den einzelnen Kliniken um die Leistungserbringung in den Kliniken und die dafür notwendigen Mittel. Es werden zwar seitens der Chefärzte auch Fragen bezüglich der Leistungsfähigkeit des Managements aufgeworfen. Bestimmte Managementmassnahmen sind aber kein Bestandteil der Verhandlungen.

Wir nehmen zusätzlich an, dass Arzt und Management wie in Abschnitt 6.4 das Budget  $c^d(\theta)$  und die freien Mittel  $t^d(\theta)$  bei gegebenem Preis  $p_1$  verhandeln, und legen damit fest, dass das Management den Preis vor der Verhandlung mit dem Arzt wählt.

### 7.6.1 Zeitliche Abfolge

Der Zeitliche Ablauf ist also wie folgt:

1. In der ersten Stufe des Spieles bestimmt der Auftraggeber die Parameter  $k(\theta)$  und  $r(\theta)$  für jede Fallschwere  $\theta \in \Theta$ .
2. In der zweiten Stufe entscheidet das Management, ob es das Vertragsangebot des Auftraggebers annehmen soll. Wenn es das Angebot annimmt, wählt das Management den Preis  $p_1$ .
3. In der dritten Stufe verhandeln das Management und der Arzt für jede Fallschwere  $\theta \in \Theta$  das Budget  $c^d(\theta)$  und die freien Mittel  $t^d(\theta)$ . Das Management entscheidet, welche Fallschweren überhaupt zur Behandlung aufgenommen werden sollen.
4. In der vierten Stufe kommen die  $n$  Patienten zur Behandlung in das Spital. Der Arzt stellt die jeweilige Fallschwere  $\theta$  der einzelnen Patienten fest und entscheidet, ob er den Patienten zur Behandlung aufnimmt oder ob er dem Patienten die Aufnahme



verweigert. Wenn er sich für die Behandlung des Patienten entscheidet, berichtet der Arzt die Fallschwere wahrheitsgemäss an das Management. Das Management entscheidet, ob es die Behandlung zulässt. Wenn das Management die Behandlung zulässt, bestimmt der Arzt die Mengen  $q_1, q_2$  zur Behandlung des Patienten so, dass das zwischen Arzt und Management verhandelte Budget  $c^d(\theta)$  eingehalten wird. Das Management berichtet die Fallschwere wahrheitsgemäss an den Auftraggeber. Die Behandlungskosten werden von allen drei Akteuren beobachtet. Die Zahlungen finden statt.

Abermals lösen wir das Spiel mittels Rückwärtsinduktion und beginnen in der vierten Stufe.

### 7.6.2 Bestimmung der Mengen $q_1$ und $q_2$ durch den Arzt

Der Arzt wählt für die Behandlung eines Patienten der Fallschwere  $\theta$  die Mengen  $q_1$  und  $q_2$  bei gegebenem Preis  $p_1$  und gegebener Budgetierung  $c^d(\theta)$ . Diese Spielstufe und ihre Ergebnisse sind also identisch mit jenen aus Abschnitt 7.5.2.

### 7.6.3 Bestimmung der Budgetierung $c^d(\theta)$ sowie der freien Mittel $t^d(\theta)$ durch Verhandlungen zwischen Arzt und Management

Management und Arzt einigen sich im Rahmen der Verhandlung für jede Fallschwere auf eine Budgetierung  $c^{d*}(\theta)$  und auf die freien Mittel  $t^{d*}(\theta)$ . Die Verhandlungen finden bei gegebenem Preis  $p_1$  statt, den das Management vor den Verhandlungen festlegt, so dass der Preis  $p_1$  an sich nicht Gegenstand der Verhandlung ist.

Allerdings stellt sich die Frage, ob das Management in den Verhandlungen darauf bestehen kann, dass die Kosten seines Arbeitsleides  $\bar{\varphi}(p_1)$  bei einer Verhandlungslösung gedeckt sein müssen. Kann das Management gegenüber dem Arzt glaubhaft darlegen, dass die Verhandlungen scheitern, wenn der Gewinn unter Berücksichtigung der Kosten des Arbeitsleides negativ ist, dann gilt für die interne Teilnahmebedingung des Managements bei der Verhandlung

$$G(\theta) = k(\theta) - r(\theta)c^d(\theta) - t^d(\theta) - \bar{\varphi}(p_1) \geq 0. \quad (139)$$

Dagegen lässt sich einwenden, dass die Anstrengungen zum Zeitpunkt der Verhandlung bereits getätigt und die Kosten dafür bereits angefallen sind. Die Kosten des Arbeitsleides lassen sich zum Zeitpunkt der Verhandlung also als versunkene Kosten auffassen und sollten streng genommen nicht mehr entscheidungsrelevant sein. In diesem Fall versucht

das Management in den Verhandlungen den „ex post“ Gewinn<sup>61</sup>  $\check{G}(\theta) = k(\theta) - r(\theta)c^d(\theta) - t^d(\theta)$  zu maximieren, und es gilt für die Teilnahmebedingung

$$\check{G}(\theta) = k(\theta) - r(\theta)c^d(\theta) - t^d(\theta) \geq 0 \quad (140)$$

Um den Einfluss dieser Problematik analysieren zu können, nehmen wir an, die Kosten des Arbeitsleides werden zu einem Teil  $0 \leq \alpha \leq 1$  in den Verhandlungen berücksichtigt, so dass wir für die interne Teilnahmebedingung des Managements bei der Verhandlung schreiben

$$\tilde{G}(\theta) = k(\theta) - r(\theta)c^d(\theta) - t^d(\theta) - \alpha\bar{\varphi}(p_1) \geq 0. \quad (141)$$

Dabei ist allerdings zu beachten, dass die Interpretation des Arbeitsleides zum Zeitpunkt der Verhandlung als (teilweise) versunkene Kosten zu einem negativen Gewinn für das Management führt, wenn der Arzt zu mächtig ist. Für den Moment sei diese Problematik ausgeblendet und es sei angenommen, der Arzt verfüge nicht über die notwendige Verhandlungsmacht. In Abschnitt 7.6.6 wird die Problematik aber nochmals aufgegriffen und kurz diskutiert.

Mit (141) lautet das Maximierungsproblem der Nash-Verhandlungslösung

$$\max_{c^d(\theta), t^d(\theta)} \pi(\theta) = (k(\theta) - r(\theta)c^d(\theta) - t^d(\theta) - \alpha\bar{\varphi}(p_1))^w (t^d(\theta) + \theta V^*(.) - u)^{1-w}. \quad (142)$$

Für die Bedingungen erster Ordnung folgt

$$\frac{w}{\tilde{G}(\theta)} = \frac{1-w}{U(\theta) - u} \quad (143)$$

und

$$-\frac{w}{\tilde{G}(\theta)}r(\theta) + \frac{1-w}{U(\theta) - u}\lambda^*(\theta, .) = 0. \quad (144)$$

Mit (143) ergibt sich für (144)

---

<sup>61</sup>Im mikroökonomischen Sprachgebrauch steht „ex post“ für den Zeitraum, nach dem die Unsicherheit bezüglich eines Umstandes aufgelöst wurde. Da im Modell die Fallschwere erst nach der Verhandlung bekannt wird, ist die Verwendung des Begriffes „ex post Gewinn“ nicht ganz korrekt. Da er aber die Situation, in der sich das Management hinsichtlich der bereits geleisteten Anstrengung befindet, sehr schön illustriert, findet er hier entgegen der üblichen Konvention Verwendung.

$$\lambda^*(\theta, c^*(\theta), p_1) - r(\theta) = 0. \quad (145)$$

Mit den Bedingungen (143) und (145) lässt sich das optimale Budget als eine Funktion der Kostenbeteiligung und neu des Preises  $p_1$  darstellen, die optimalen freien Mittel sind dagegen eine Funktion der Kostenbeteiligung, der internen Machtverhältnisse, des Preises  $p_1$  und des Parameters  $\alpha$ :

$$c^{d*}(\theta, \cdot) \equiv c^{d*}(\theta, r(\theta), p_1). \quad (146)$$

$$t^{d*}(\theta, \cdot) \equiv t^{d*}(\theta, r(\theta), p_1, w, \alpha) \quad (147)$$

#### 7.6.4 Bestimmung des Preises $p_1$ durch das Management

Der „ex post“ Nutzenüberschuss von Arzt und Management bei einem Patienten der Fall-schwere  $\theta$  zum Zeitpunkt der Verhandlung ist gegeben mit

$$\tilde{N}\ddot{U}(\theta) = \tilde{G}(\theta) + U(\theta) - u. \quad (148)$$

Für den Anteil des Arztes am gemeinsamen Nutzenüberschuss ergibt sich mit (143)

$$\frac{U(\theta) - u}{\tilde{N}\ddot{U}(\theta)} = 1 - w. \quad (149)$$

Damit gilt für den Nutzen des Arztes

$$U(\theta) = t^{d*}(\theta, \cdot) + \theta V^*(\cdot) = (1 - w)(k(\theta) - r(\theta)c^{d*}(\theta, r(\theta), p_1) + \theta V^*(\cdot) - \alpha \bar{\varphi}(p_1)) + u. \quad (150)$$

Für die freien Mittel folgt

$$t^{d*}(\theta, \cdot) = (1 - w)(k(\theta) - r(\theta)c^{d*}(\theta, r(\theta), p_1) - \alpha \bar{\varphi}(p_1)) - w\theta V^*(\cdot) + u. \quad (151)$$

Zum Zeitpunkt, zu dem das Management den Preis  $p_1$  bestimmt, kennt es das Angebot des Auftraggebers, die Verhandlungen mit dem Arzt stehen aber erst noch bevor. Bei seiner Entscheidung, wie intensiv es die Bemühungen zur Kostensenkung verfolgen soll, berücksichtigt das Management neben dem dafür anfallenden Arbeitsleid auch den Einfluss seiner Entscheidung auf seinen Anteil aus dem zukünftigen Verhandlungsergebnis,

d.h. seinen Anteil aus dem gemeinsamen „ex post“ Nutzenüberschuss. Da die freien Mittel für den Arzt und damit auch der Anteil des Nutzenüberschusses für das Management davon abhängen, welcher Teil  $\alpha$  der Kosten aus dem Arbeitsleid in den Verhandlungen berücksichtigt werden, muss das Management zumindest Erwartungen hinsichtlich  $\alpha$  bilden, um seinen Anstrengungen bzw. den Preis  $p_1$  festzulegen. Wir nehmen an, das Management könne  $\alpha$  zum Zeitpunkt, an dem es die Anstrengungen leistet, korrekt antizipieren. Damit lautet das Maximierungsproblem des Managements zur Bestimmung des optimalen Preises  $p_1$

$$\max_{p_1} E[G(\theta)] = E[k(\theta) - r(\theta)c^{d*}(\theta, \cdot) - t^{d*}(\theta, r(\theta), p_1, w, \alpha)] - \bar{\varphi}(p_1). \quad (152)$$

Mit (151) ergibt sich für das Maximierungsproblem

$$\max_{p_1} E[G(\theta)] = w(E[k(\theta) - r(\theta)c^{d*}(\theta, \cdot) + \theta V^*(\cdot)]) - (1 - (1 - w)\alpha)\bar{\varphi}(p_1) - u. \quad (153)$$

Wenn das Management das Verhandlungsergebnis korrekt antizipiert, dann folgt für die Bedingung erster Ordnung für den optimalen Preis  $p_1$  mit (145)

$$-wE[r(\theta)q_1^*(\cdot)] - (1 - (1 - w)\alpha)\bar{\varphi}'(p_1^*) = 0. \quad (154)$$

Wenn das Arbeitsleid des Managements in den späteren Verhandlungen vollständig berücksichtigt wird ( $\alpha = 1$ ), dann entspricht die Bedingung (154) der Bedingung erster Ordnung für den optimalen Preis unter der PA-Lösung (137). In diesem Fall sind die beiden Lösungen äquivalent, d.h. das Arbeitsleid wird zwischen Management und Arzt vollständig internalisiert und der gesamte spitalinterne Entscheidungsprozess ist zwischen Arzt und Management Pareto-effizient. Insbesondere ist der Preis unabhängig von der Verteilung der Verhandlungsmacht,  $\frac{dp_1^*}{dw} = 0$ .

Wird das Arbeitsleid dagegen nur zu einem Teil oder gar nicht in den späteren Verhandlungen berücksichtigt ( $\alpha \leq 1$ ), gilt  $(1 - (1 - w)\alpha) > w$  und das Management gewichtet das Arbeitsleid höher als die Kostenersparnis. In diesem Fall bestimmt das Management unter der Verhandlungslösung einen höheren Preis als bei der PA-Lösung, da es vom positiven Effekt eines tieferen Preises auf den gemeinsamen Nutzenüberschuss nur zu einem Teil profitiert. Dieser Anteil wird umso geringer, je geringer die Verhandlungsmacht des Managements gegenüber dem Arzt ist. Mit impliziter Differenzierung von (154) folgt

$$\frac{dp_1^*}{dw} = \frac{E[r(\theta)q_1^*(\cdot)] - \alpha\bar{\varphi}'(p_1^*)}{-wE[r(\theta)(\partial q_1^*(\cdot)/\partial p_1 + \partial q_1^*(\cdot)/\partial c^*(\cdot)\partial c^*(\theta, \cdot)/\partial p_1)] - (1 - (1 - w)\alpha)\bar{\varphi}''(p_1^*)} < 0$$

wobei der Ausdruck bei erfüllter Bedingung zweiter Ordnung des Maximierungsproblems (153) kleiner null ist (die Bedingung zweiter Ordnung steht im Nenner von (155)). Je grösser die Verhandlungsmacht des Arztes gegenüber dem Management ist, umso weniger Anstrengungen unternimmt das Management und umso höher ist der resultierende Preis  $p_1$ . Da die Anstrengungen des Managements in den Verhandlungen nicht oder nicht vollständig internalisiert werden, ist die NV-Lösung nicht Pareto-effizient.

**Ergebnis 7.2** *Wenn das Arbeitsleid in den Verhandlungen vollständig berücksichtigt wird ( $\alpha=1$ ), wählt das Management denselben Preis wie unter der PA-Lösung, unabhängig von der Verteilung der Verhandlungsmacht. In diesem Fall resultiert zwischen Arzt und Management eine Pareto-effiziente Lösung und die NV-Lösung ist mit der PA-Lösung äquivalent. Wird das Arbeitsleid dagegen nur zu einem Teil oder gar nicht berücksichtigt ( $\alpha \leq 1$ ), profitiert das Management nur in dem Mass von einem tieferen Preis, mit dem es seine Interessen in der Verhandlung mit dem Arzt durchsetzen kann. Daher lohnen sich die Anstrengungen einen niedrigeren Preis zu erreichen umso weniger, je weniger Verhandlungsmacht das Management gegenüber dem Arzt hat. Bei gegebener Kostenbeteiligung ist dann eine grössere Verhandlungsmacht des Arztes mit einem höheren Preis  $p_1$  verbunden und die NV-Lösung ist zwischen Arzt und Management nicht Pareto-effizient.*

### 7.6.5 Maximierungsproblem des Auftraggebers

Für  $\alpha = 1$  sind PA-Lösung und NV-Lösung äquivalent, so dass auch das Maximierungsproblem des Auftraggebers dasselbe ist wie bei spitalinternem PA-Verhältnis zwischen Management und Arzt. Auch für  $\alpha \leq 1$  resultiert grundsätzlich dasselbe Maximierungsproblem für den Auftraggeber, allerdings hat es der Auftraggeber hier mit einem Management zu tun, welches ein umso grösseres Gewicht auf das Arbeitsleid legt, je geringer seine Verhandlungsmacht gegenüber dem Arzt ist und daher umso „träger“ auf die Anreize der Kostenbeteiligung reagiert.

Es folgt, dass bei der NV-Lösung verglichen mit der PA-Lösung ein stärkerer Anreiz und deshalb insgesamt eine höhere Kostenbeteiligung notwendig ist, um denselben Preis  $p_1$  zu erreichen, wenn  $\alpha \leq 1$  der Fall ist. Sei zum Beispiel  $\alpha = 0$ . Aus einem Vergleich der Bedingungen (154) und (127) folgt, dass das Management bei *gegebenen Mengen*  $q_1$  und  $q_2$  nur dann den effizienten Preis wählt, wenn  $wE[r(\theta)q_1^*(.)] = E[q_1^*(.)]$  gilt, was zum Beispiel bei einer konstanten Beteiligung über alle Fallschweren  $r = 1/w > 1$  erfüllt ist. Bei der PA-Lösung ist dies dagegen bei  $r = 1$  der Fall.

**Ergebnis 7.3** *Wenn das Arbeitsleid des Managements zum Zeitpunkt der Verhandlungen von Arzt und Management als (teilweise) versunkene Kosten gesehen wird, dann ist die Äquivalenz zwischen der PA-Lösung und der NV-Lösung nicht gegeben. Um einen bestimmten Preis  $p_1$  zu erreichen, sind dann bei der NV-Lösung stärkere Anreize notwendig als bei der PA-Lösung.*

Da sich auch unterschiedliche bestmögliche Allokationen aus PA-Lösung und NV-Lösung ergeben, wenn  $\alpha \leq 1$  gilt, bewertet der Auftraggeber auch beide Lösungen unterschiedlich. Ohne weitere Annahmen ist es zwar nicht eindeutig, ob der Auftraggeber die PA-Lösung oder die NV-Lösung vorzieht, intuitiv wird es aber eher die PA-Lösung sein, da diese als Pareto-effiziente Lösung für jedes Budget die Vergütung des Spitals minimiert und einen niedrigeren Preis  $p_1$  beinhaltet.

Erwähnenswert ist in diesem Zusammenhang zum einen noch, dass im Vergleich zur PA-Lösung bei der Verhandlungslösung nicht etwa die Präferenzen des Arztes, sondern ein Teil der Präferenzen des Managements - das Arbeitsleid - umso mehr gewichtet wird, je grösser die Verhandlungsmacht des Arztes ist. Zum anderen stellt sich bei der NV-Lösung der aus Sicht des Auftraggebers effiziente Preis bzw. die effiziente Intensität der Massnahmen nicht bei einem Vergütungssystem ein, welches dem Spital die Kostenverantwortung zum 100% zuteilt. Um diesen Preis zu erreichen, ist eine Kostenbeteiligung von über 100% notwendig, d.h. das Spital müsste für die marginalen Kosten noch einen Zuschlag zahlen.

#### 7.6.6 Die Kosten des Arbeitsleides und negativer Gewinn des Managements

Die Interpretation des Arbeitsleides (teilweise) als versunkene Kosten zum Zeitpunkt der Verhandlung verlangt eine Diskussion der internen sowie der externen Teilnahmebedingung des Managements.

Wird mit  $\alpha = 0$  das Arbeitsleid zum Beispiel vollständig als versunkene Kosten betrachtet und hat daher keinen Einfluss auf die Verhandlungen, folgt für die Teilnahmebedingung des Managements während der Verhandlungen

$$\tilde{G}(\theta) = k(\theta) - r(\theta)c(\theta) - t^d(\theta) \geq 0. \quad (156)$$

Dies birgt das Problem, dass für das Management im Endeffekt ein negativer Gewinn (ein Gewinn unter dem internen Reservationsgewinn also) resultieren kann. Da der interne Reservationsgewinn in diesem Fall lediglich verlangt, dass  $k(\theta) - r(\theta)c(\theta) - t^d(\theta) > 0$  erfüllt ist, sind Situationen möglich, in denen der tatsächliche Gewinn

$$G(\theta) = k(\theta) - r(\theta)c(\theta) - t^d(\theta) - \bar{\varphi}(p_1) \quad (157)$$

negativ ist. Das Problem ist vor allem dann in Betracht zu ziehen, wenn die Verhandlungsmacht des Arztes hoch ist.

Wenn das Management antizipiert, dass der Anteil des Nutzenüberschusses, der ihm nach der Verhandlung bleibt, nicht ausreicht, um die Anstrengungen zu finanzieren, wird es dem Angebot des Auftraggebers in der vorigen Stufe erst gar nicht zustimmen, selbst wenn das Angebot die bisher gültige Bedingung  $N\ddot{U}(\theta) = g$  erfüllt. Insbesondere, wenn die spitalinternen Machtverhältnisse zugunsten des Arztes sind, ist für den Auftraggeber eine weitere Teilnahmebedingung zu beachten, die dem Management einen positiven Gewinn sichert. Diese erreicht der Auftraggeber mit einer Erhöhung der Pauschale  $k(\theta)$ . Da jedoch jede zusätzliche Geldeinheit, die der Auftraggeber dem Spital vergütet, zwischen Arzt und Management aufgeteilt wird und nicht vollumfänglich dafür verwendet werden kann, den Mindestgewinn für das Management zu sichern, kommt die Mindestgewinnbedingung dem Auftraggeber unter Umständen sehr teuer zu stehen.

## 7.7 Fazit Kapitel 7

Durch die zusätzliche Informationsasymmetrie hinsichtlich der Anstrengungen des Managements resultiert für den Auftraggeber ein Trade-off zwischen interner Budgetierung bzw. den Behandlungsmengen und dem Preis  $p_1$ , unabhängig davon, ob nun die PA-Lösung relevant ist oder die NV-Lösung. Mit der Festlegung der Beteiligung für jede Fallschwere  $r(\theta) \forall \theta \in \Theta$  nimmt der Auftraggeber Einfluss auf die interne Budgetierung  $c^d(\theta)$  sowie auf die Anstrengungen des Managements bzw. den Preis  $p_1$ . Dabei reicht die bei gegebenem Preis  $p_1$  optimale Beteiligung aus Kapitel 6 nicht aus, um bei nun *endogenem* Preis dem Management Anreize zu geben, einen effizienten Preis zu bestimmen.

Hinsichtlich der Bedeutung des spitalinternen Mechanismus ist das entscheidende Ergebnis der Analyse in diesem Kapitel, dass sich die Lösungen des PA-Mechanismus und des NV-Mechanismus unterscheiden können, die Gründe für die Differenz jedoch nicht direkt aus der alleinigen Tatsache folgen, dass in einem Fall die Parameter vom Management festgelegt werden und im anderen von Management und Arzt. Die Unterschiede resultieren vielmehr aus verschiedenen Konstellationen im gesamten Entscheidungsprozess, die bei der NV-Lösung plausibel sind, bei der PA-Lösung jedoch nicht. Wie oben dargelegt, ist es durchaus denkbar, dass das vom Management bereits geleistete Arbeitsleid in den Verhandlungen über die freien Mittel und die Budgetierung nicht oder nur teilweise Bedeutung hat. Bei der PA-Lösung wäre diese Annahme dagegen nicht zu begründen und vollkommen unplausibel, da das Management dort die freien Mittel und die Budgetierung autonom bestimmt und dabei auf jeden Fall dafür sorgen wird, dass sein angefallenes Ar-

beitsleid finanziert wird. Der Unterschied der NV-Lösung zur PA-Lösung tritt also nicht als Konsequenz der Verhandlungen an sich auf, die ja für sich genommen Pareto-effizient sind, sondern entsteht durch die mit der Verhandlung verbundenen Rahmenbedingungen.

Für den Auftraggeber bedeutet dies, dass er nicht nur darüber informiert sein muss, ob die freien Mittel und die Budgetierung durch das Management oder im Rahmen von Verhandlungen zwischen Management und Arzt bestimmt werden, sondern dass er auch über Informationen hinsichtlich weiterer Umstände und Aspekte, wie zum Beispiel den Anteil  $\alpha$  oder die Verteilung der Verhandlungsmacht, verfügen muss.



## 8 Asymmetrische Information hinsichtlich der Fallschwere

In den beiden vorhergehenden Kapiteln 6 und 7 fand die Analyse des Modells mit doppeltem Unterstellungsverhältnis unter der Annahme statt, der Arzt berichte die Fallschwere des aktuell behandelten Patienten wahrheitsgemäss an das Management und das Management seinerseits leite den Bericht des Arztes unverfälscht an den Auftraggeber weiter. Von dieser Annahme sehen wir nun ab. Stattdessen gehen wir in diesem Kapitel davon aus, Arzt bzw. Management wählen die berichtete Fallschwere nach dem Kriterium der Nutzen- bzw. Gewinnmaximierung. Im Falle des Arztes nehmen wir also an, sein Bericht an das Management entspricht nur dann der von ihm ermittelten Fallschwere, wenn der wahrheitsgemässe Bericht seinen Nutzen maximiert. Beim Management unterstellen wir entsprechend, der Bericht an den Auftraggeber enthalte nur dann die vom Arzt berichtete Fallschwere, wenn die unverfälschte Weiterleitung der Information den Gewinn des Managements maximiert.

Mit der asymmetrischen Information bezüglich der exogen gegebenen Fallschwere greifen wir die Problematik der versteckten Information auf. Der Fokus dieses Kapitels liegt auf der Frage, wie sich die asymmetrische Information an sich auf die optimal erreichbare Allokation unter spitalinternem PA-Mechanismus auswirkt. Insbesondere soll analysiert werden, ob die spitalinterne Informationsverteilung für den Auftraggeber von Relevanz ist. Hierzu vergleichen wir zwei hinsichtlich der Informationsverteilung unterschiedliche Szenarien. Im ersten Szenario besteht lediglich zwischen Auftraggeber und Management asymmetrische Information hinsichtlich der Fallschwere. Spitalintern nehmen wir an, der Arzt berichte die Fallschwere immer wahrheitsgemäss (zum Beispiel, da das Management über geeignete Kontrollmechanismen verfügt.) Im zweiten Szenario kommt es zusätzlich zwischen Management und Arzt zu asymmetrischer Information über die Fallschwere.

Um den Einfluss der asymmetrischen Verteilung der Information an sich herauszuarbeiten, schliessen wir verzerrende Effekte, die ihre Ursache in der Vertragsausgestaltung haben, aus, indem wir bei der Analyse annehmen, Auftraggeber und Management setzen jeweils einen „optimalen“ Vertrag ein. Es wird sich zeigen, dass der unter den optimalen Verträgen in jedem Fall das direkte Vergütungs- und Budgetierungssystem  $t(\theta), c(\theta)$  ist, so dass wir insbesondere den Auftraggeber nicht von vorneherein auf das lineare System  $t^m(\theta) = k(\theta) - r(\theta)c$  festlegen.

Die Struktur dieses Kapitels ist wie folgt: In Abschnitt 8.1 betrachten wir den Fall, dass es lediglich zwischen Auftraggeber und Management zu asymmetrischer Information hin-

sichtlich der Fallschwere kommt. In Abschnitt 8.2 betrachten wir dagegen die Situation, in der auch zwischen Arzt und Management asymmetrische Information hinsichtlich der Fallschwere besteht.

In Abschnitt 8.3 wird untersucht, welche Bedeutung es bei asymmetrischer Information über die Fallschwere hat, wenn es für Arzt und Management unattraktiv ist, Patienten nach ihrer Fallschwere zu selektieren und ihnen gegebenenfalls die Behandlung zu verweigern. In Abschnitt 8.4 am Schluss dieses Kapitels wird kurz auf den Einfluss der spitalinterne Verhandlungslösung auf das Verhalten des Spitals eingegangen.

## **8.1 Asymmetrische Information hinsichtlich der Fallschwere zwischen Management und Auftraggeber**

### **8.1.1 Hintergrund**

**8.1.1.1 Diagnosis Related Groups (DRG) - Abhängigkeit der Vergütung von der Fallschwere** In den letzten Jahrzehnten haben sich so genannte DRG - Abrechnungssysteme als Grundlage der Vergütung von stationären Spitalleistungen in vielen Ländern etabliert. Erstmals 1983 in den USA zur Vergütung der stationären Spitalbehandlung von Medicare Patienten eingesetzt, kommen DRG-Systeme gegenwärtig in den meisten europäischen Ländern zur Anwendung [Schreyögg (2006)]. Auch in der Schweiz wird ab 2012 hauptsächlich über ein DRG-System abgerechnet. Im Detail gibt es zwischen den Systemen von Land zu Land Unterschiede, die grundlegende Systematik ist jedoch bei allen DRG-Systemen dieselbe: Im Wesentlichen besteht ein DRG-System aus mehreren hundert „Diagnosegruppen“, den so genannten DRG's (in der Schweiz sind es im Einführungsjahr 2012 ca. 1'100 DRG's). Jeder stationäre Behandlungsfall wird unter einem DRG-System nach der Behandlung genau einer DRG zugeordnet. Hauptkriterium der Zuordnung ist die so genannte „Hauptdiagnose“, der eigentliche Grund, weshalb der Patient im Spital behandelt wird. Weitere Kriterien sind z.B. durchgeführte spezielle medizinische Leistungen sowie persönliche Eigenschaften des Patienten wie Alter oder Geschlecht, die sich auf den Pflege- und Behandlungsaufwand auswirken.

Jede DRG ist wiederum mit einer dimensionslosen Relativzahl, meist als „Kostengewicht“ bezeichnet, verknüpft. Das Kostengewicht lässt sich als eine zahlenmässige Darstellung der Fallschwere interpretieren. Es spiegelt kurz gesagt den durchschnittlichen Behandlungsaufwand eines Patienten dieser DRG im Verhältnis zu den durchschnittlichen Behandlungskosten aller stationären Fälle wider. Die Höhe der Vergütung, die ein Spital unter einem DRG-System für die Behandlung eines Patienten einer bestimmten DRG erhält, ergibt sich aus der Multiplikation des Kostengewichtes mit einem Basispreis (der so ge-

nannten Baserate).<sup>62</sup> Mit der Interpretation des Kostengewichts als Indikator der Schwere der Erkrankung eines Patienten ist unter einem DRG-System die Höhe der Vergütung also eine Funktion der Fallschwere.

**8.1.1.2 Bestimmung der Fallschwere im Spital** Der Prozess der Zuordnung der Fallschwere zu einem Patienten unter einem DRG-System lässt sich wie folgt grob darstellen: Die Diagnosen und die zur Behandlung des Patienten durchgeführten Leistungen werden vom klinischen Personal, vor allem von den Ärzten, in der Patientenakte dokumentiert. Nach Entlassung des Patienten aus dem Spital gelangt die Patientenakte in eine „Kodierabteilung“, die in der Regel der Spitaldirektion angegliedert ist. Dort werden die in der Patientenakte dokumentierten Diagnosen und medizinischen Leistungen von spezialisierten Fachkräften anhand von Klassifikationssystemen verschlüsselt in Codes dargestellt. Für die Diagnosen ist dies in der Schweiz (und in den meisten anderen Ländern, in denen ein DRG-System Anwendung findet) das ICD-10 Klassifikationssystem, für Operationen und andere medizinische Prozeduren ist es der CHOP - Katalog (Schweizerische Operationsklassifikation). Die Verschlüsselung der in der Patientenakte dokumentierten Diagnosen und Leistungen wird oftmals als „Kodieren“ bezeichnet.

Anhand der Diagnosen- und medizinischen Prozedurenkodes sowie weiteren Patientenmerkmalen, erfolgt dann eine Zuordnung des Patienten zu einer DRG. Wie bereits erwähnt ist hier vor allem die Hauptdiagnose von Bedeutung. Der Algorithmus der Zuordnung ist für alle Spitäler identisch und wird von einer speziellen Software, dem so genannten „Grouper“ ausgeführt.

Für die Ermittlung des Kostengewichts bzw. der Fallschwere sind also die Dokumentation der Diagnosen und medizinischen Prozeduren, die Kodierung derselben sowie die Zuordnung zu einer DRG die entscheidenden Vorgänge. Während die Zuordnung des Patienten zu einer DRG einem festen Algorithmus folgt und dieser letzte Schritt vom Spital nicht zu beeinflussen ist, lassen die Dokumentation der Diagnosen und medizinischen Prozeduren sowie die Kodierung einige Spielräume für das Spital zu [vgl. Dafny (2005)]. Beide Vorgänge sind zum einen recht komplex und finden so oft statt wie das Spital stationäre Fälle hat. Da beide Vorgänge von Spitalmitarbeitern intern durchgeführt werden, sind sie von Aussenstehenden kaum zu beobachten und nur in Einzelfällen zu überprüfen. Dies vor Augen halten wir also fest:

**Folgerung 8.1** *Der Prozess, der zur Feststellung der Fallschwere führt, findet weitestgehend im Spital statt und ist von Aussenstehenden kaum zu beobachten. Das Spital hat*

---

<sup>62</sup>Die Bestimmung der Baserate wiederum wird unterschiedlich gehandhabt. In der Schweiz beispielsweise bilden die Fallkosten die Basis für Verhandlungen des Preises zwischen Spitälern und Krankenkassen (für andere europäische Länder vgl. Schreyögg (2006)).

*bei der Darstellung der Fallschwere also einige Spielräume, so dass die tatsächliche Fallschwere weitestgehend als private Information des Spitals angesehen werden kann.*

**8.1.1.3 Upcoding und Kontrolle** Wenn das Spital nun gewisse Spielräume bei der Darstellung der Fallschwere hat, dann ist zu vermuten, dass das Spital seine Spielräume bei der Darstellung der Fallschwere auch zu seinem Vorteil ausnutzt, insbesondere, wenn sein Vergütung wie bei einem DRG-System von der dargestellten Fallschwere abhängt.

Vor allem aus den USA gibt es empirische Belege, dass Spitäler dies tatsächlich tun [Dafny (2005), Silverman (2004)]. In diesem Zusammenhang ist in der Praxis eine unter dem Begriff „Upcoding“ bekannte Problematik immer wieder in der Diskussion [vgl. Fiori (2010)]. Ein Spital praktiziert Upcoding, wenn es durch bewusste Manipulation an Diagnose- bzw. Operationscodes für einen Patienten eine DRG mit einem höheren Kostengewicht erwirkt, die eigentlich nicht gerechtfertigt ist. Häufig erfolgt Upcoding im Zusammenhang mit der Festlegung der Hauptdiagnose. Im Normalfall werden bei einem Behandlungsfall mehrere Diagnosen gestellt, wovon eine als Hauptdiagnose bestimmt wird. Die Hauptdiagnose ist dabei massgeblich für die Zuordnung eines Behandlungsfalles zu einer DRG [Strüwe (2011)] und damit entscheidend für die Höhe der Vergütung. Bei der Entscheidung, welche der Diagnosen als Hauptdiagnose geführt wird, sollte das Spital vor allem medizinische Kriterien beachten. Allerdings kann das Spital den Patienten auch „upcoden“, indem es eine Diagnose als Hauptdiagnose festlegt, die zu einer höheren Vergütung führt, auch wenn eine Festlegung nach den tatsächlichen medizinischen Umständen in einer anderen Hauptdiagnose und einer niedrigeren Vergütung resultieren würde.<sup>63</sup>

Um dem Problem des Upcodings zu begegnen, sind in der Praxis Einzelfallprüfungen vorgesehen. In Deutschland beispielsweise wird der Grossteil der Prüfungen vom „Medizinischen Dienst der Krankenversicherung“ (MDK), ein von den Kassen finanzierter Beratungs- und Begutachtungsdienst, durchgeführt. Insgesamt werden ca. 11% der Krankenhausabrechnungen überprüft, allerdings haben die Prüfungen nur in weniger als 23% der Kontrollen Fragen von Fehlkodierungen zum Inhalt [Fiori (2010)].

In der Schweiz sollen Bestrebungen zum Upcoding zukünftig über die so genannte „Kodierrevision“ verhindert werden. Vorgesehen ist, dass ein unabhängiger Revisor in einem Spital die ICD-10 und CHOP Kodierung in 100 - 300 Fällen pro Jahr überprüft.<sup>64</sup> Hierzu kodiert der Revisor die Fälle der Stichprobe anhand der vollständigen Patientendossiers

---

<sup>63</sup>Bei diesem Verhalten muss es sich nicht unbedingt um Betrug im strafrechtlichen Sinne handeln. Betrug würde vorliegen, wenn bewusst falsche Diagnosen eingesetzt werden. Vielmehr existiert oftmals ein Graubereich, den das Spital bei einer „kreativen“ Interpretation evtl. auszunutzen versucht.

<sup>64</sup>vgl. „Regeln und Definitionen zur Fallabrechnung unter SwissDRG“ auf der SwissDRG Homepage.

und vergleicht seine Ergebnisse mit jenen des Spitals. Das Ergebnis seiner Überprüfung wird dann dem Spital, den Kassen und dem Kanton zugänglich gemacht.

Allerdings sind solche Revisionen ebenfalls mit Ineffizienzen verbunden, da Einzelfallprüfungen sehr teuer sind<sup>65</sup>, aber keinen direkten Nutzen enthalten. Bei der Durchführung bzw. der Häufigkeit der Kontrollen muss also zwischen den Effizienzgewinnen durch verhindertes Upcoding und durch die Kosten der Kontrolle resultierenden Ineffizienzen abgewogen werden.

Eine theoretische Arbeit zu diesem Thema liegt von Kuhn/Siciliani (2008) vor. Die Autoren zeigen, dass aus Sicht des Auftraggebers der optimale Preis für ein Kostengewicht unter dem first best-Preis liegt, wenn das Spital das Kostengewicht „upcoden“ kann. Der marginale Gewinn durch Manipulation steigt mit dem Preis und damit auch das Ausmass der Manipulation. Von daher ist es für den Auftraggeber besser, den Preis unter das first best-Niveau zu setzen, um dem Spital weniger Anreize zum Upcoding zu bieten. Mit der Möglichkeit der stichprobenartigen Kontrolle (bei gegebener Sanktion) kann der Auftraggeber die Verzerrung verringern, indem er mit einer Erhöhung der Häufigkeit der Kontrollen einen höheren optimalen Preis erreichen kann. Da die Kontrollen aber selbst Kosten verursachen, sucht der Auftraggeber den optimalen Trade-off zwischen Verzerrung durch Upcoding und den Kosten der Kontrolle. Es resultiert ein Preis, der zwar etwas höher ist also ohne das Instrument der Kontrolle, aber immer noch unterhalb des first best-Preises.

Rein theoretisch könnte man den Trade-off so gut wie beseitigen, indem man die Häufigkeit der Kontrollen äusserst gering hält, die Strafe bei aufgedecktem Upcoding aber dabei so hoch ansetzt, dass das Spital trotz geringer Wahrscheinlichkeit der Entdeckung vom Upcoding absieht. In der Realität ist die Höhe der Bestrafung jedoch begrenzt, zum Beispiel deshalb, weil der Kodierungsprozess häufig ein sehr komplexer Vorgang und ein unbeabsichtigtes Upcoding seitens des Spitals möglich ist. Drakonische Strafen wären darum politisch kaum durchzusetzen und aus rechtlicher Sicht wohl auch problematisch. Wenn das Spital also an sich die Bereitschaft hat, ökonomische Aspekte in die Kodierung einfließen zu lassen, dann lässt sich Upcoding durch eine stichprobenartige Kodierrevision vielleicht mindern. Ausschliessen lässt es sich aufgrund der begrenzten „Hebelwirkung“ jedoch nicht. Hinzu kommt, dass die durchgeführten Kontrollen nur bestimmte Formen des Upcodings aufdecken können. So mag vielleicht eine fehlerhafte Kodierung der aus dem Patientendossier entnommenen Informationen bemerkt werden können. Fliessen aber schon bei der

---

<sup>65</sup>vgl. Stellungnahmen des Verbands der Universitätsklinika Deutschlands zu Fragen des deutschen Bundesministeriums für Gesundheit, 2007, Seite 4, erhältlich auf der Homepage des VUD (<http://www.uniklinika.de>)

Erstellung des Dossiers ökonomische Überlegungen ein, sind diese kaum aufzudecken.

### 8.1.2 Anreiztheoretischer Ansatz: Prinzip der direkten Offenbarung und Anreize zur wahrheitsgemässen Berichterstattung

Aus den Ausführungen des obigen Abschnittes lässt sich folgern, dass die Kontrollen der Berichterstattung ein notwendiges Übel sind, das umso notwendiger wird, je grösser der Schaden bzw. die Ineffizienzen aus der falschen Berichterstattung ist bzw. sind. Ist der Schaden sehr gross, müssen auch die Häufigkeit und Intensität der Kontrollen sehr ausgeprägt sein.

Wie gross der Schaden aus der falschen Berichterstattung ist, hängt allerdings zu einem grossen Teil von der Ausgestaltung des Vergütungssystems ab. Angesichts der durch die Kontrollen selbst verursachten Ineffizienzen kann man daher die Frage stellen, ob es für einen Auftraggeber nicht sinnvoller ist, durch kluge Ausgestaltung eines Vergütungssystems, in dem das Verhalten des Spital hinsichtlich seiner Berichterstattung bereits berücksichtigt ist, die Ineffizienzen von vorneherein zu minimieren und *ohne* die Kontrollen auszukommen.

In der bereits in Kapitel 3 erwähnten Anreiztheorie ist die asymmetrische Information hinsichtlich der Fallschwere mit dem Problem der nicht wahrheitsgemässen Darstellung durch das Spital der Problematik der *versteckten Information* zuzuordnen. Der sogenannte ‚Mechanism Design‘-Ansatz in der Anreiztheorie besteht in der Ausgestaltung eines optimalen Vergütungssystems, welches das Verhalten des Agenten (Spital) bei versteckter Information in optimaler Weise koordiniert. In ‚optimaler Weise‘ bedeutet in dem hier diskutierten Zusammenhang, dass durch das optimale Vergütungssystem die aus der Problematik der versteckten Informationen hinsichtlich der Fallschwere resultierenden Ineffizienzen minimiert werden und die unter den gegebenen Umständen bestmögliche Allokation resultiert.

Der mit dieser Fragestellung verbundene Optimierungsprozess ist allerdings zunächst alles andere als trivial. Denn hierzu muss natürlich bekannt sein, *welches* System an sich überhaupt in Frage kommt, um die bestmögliche Allokation zu erreichen. So ist es zum Beispiel möglich, dass ein System mit linearer Kostenbeteiligung, wie es in den Kapiteln 6 und 7 zur Anwendung kam, hier nicht die bestmögliche Allokation erreicht. Um dagegen des Vergütungssystem, welches die bestmögliche Allokation erreicht, zu bestimmen, müsste man im Prinzip für jedes denkbare Vergütungssystem die optimale Allokation ermitteln und mit den optimalen Allokationen aller anderen System vergleichen [vgl. Breyer (2005)].

Diese Unterfangen wird durch das so genannte „Prinzip der direkten Offenbarung“ erheblich erleichtert. In der Anwendung im Rahmen des Modells mit doppeltem Unterstellungsverhältnis besagt das Prinzip der direkten Offenbarung, dass jede Allokation, die von einem beliebigen Vergütungssystem erreicht wird, auch mit einem Vergütungssystem möglich ist, das direkt nach der Ausprägung des Parameters (in dem hier diskutierten Zusammenhang also die Fallschwere) fragt und für jede Ausprägung eine Vergütung  $t^d(\theta)$  und eine Kostenvorgabe  $c^m(\theta)$  definiert.

Wenn wir also nach der bestmöglichen Allokation fragen, die gegeben der asymmetrischen Information über die Fallschwere überhaupt möglich ist, erspart uns das Prinzip der direkten Offenbarung, alle möglichen Vergütungssysteme nach ihrer optimalen Allokation abzusuchen und untereinander zu vergleichen. Es reicht, wenn wir das direkte Vergütungssystem betrachten und die damit optimal erreichbare Allokation ermitteln. Diese Allokation ist dann auch die bestmögliche Allokation überhaupt. Eventuell erreicht auch ein anderes Vergütungssystem dieselbe Allokation, eine „bessere“ ist aber nicht möglich.

Zur Illustration des Prinzips im Rahmen des Modells mit doppeltem Unterstellungsverhältnis<sup>66</sup> stellen wir uns die Situation aus Kapitel 6 vor, allerdings mit dem Unterschied, dass das Management die Fallschwere nur dann korrekt an den Auftraggeber berichtet, wenn die wahrheitsgemässe Berichterstattung für das Management optimal ist. Der Arzt berichte die Fallschwere weiterhin in jedem Fall wahrheitsgemäss an das Management. Wenn der Auftraggeber wie in Kapitel 6 zur Vergütung des Managements die Parameter  $k(\theta)$  und  $r(\theta)$  des cost-sharing Systems  $t^m(\theta, c) = k(\theta) - r(\theta)c$  von der vom Management berichteten Fallschwere abhängig macht, kann das Management mit dem Bericht der Fallschwere die Vergütung beeinflussen. Wir haben gesehen, dass das Management in Kapitel 6 dem Arzt gerade den Reservationsnutzen  $u$  gewährt und im Endeffekt für jede Fallschwere  $\theta \in \Theta$  die Funktion

$$G(\theta) = k(\theta) - r(\theta)c^d(\theta) + \theta V^*(c^d(\theta)) - u \quad (158)$$

mit der Wahl der optimalen Budgetierung  $c^d(\theta)$  maximiert, wobei  $V^*(c^d(\theta)) \equiv V(q_1^*(c^d(\theta)), q_2^*(c^d(\theta)))$  mit  $q_1^*(c^d(\theta))$  und  $q_2^*(c^d(\theta))$  als die vom Arzt gewählten Mengen bei der Budgetierung  $c^d(\theta)$  sind. Nun hat das Management zusätzlich zur Wahl der Budgetierung  $c^d(\theta)$  die Möglichkeit, auch mit dem Bericht der Fallschwere seinen Gewinn zu maximieren.

---

<sup>66</sup>Der formelle Beweis des Prinzips in seiner allgemeinen Formulierung ist an vielen Stellen zu finden, z.B. Laffont/Tirole (1993), S. 120 oder Mas-Colell/Whinston/Green (1995), S. 871, so dass hier auf seine Darstellung verzichtet wird.

Sei  $\hat{\theta}^m(\theta)$  der Bericht der Fallschwere bei tatsächlicher Fallschwere  $\theta$ , dann bestimmt das Management also das Tupel  $\{c^d(\theta), \hat{\theta}^m(\theta)\}$  zur Maximierung seines Gewinns. Im spieltheoretischen Sinne stellt die Bestimmung von  $\{c^d(\theta), \hat{\theta}^m(\theta)\}$  die Festlegung der Strategie des Managements bei einem Patienten der Fallschwere  $\theta$  dar. Für alle Fallschweren  $\theta \in \Theta$  sei  $\{c^{d*}(\theta), \hat{\theta}^{m*}(\theta)\}$  die optimale Strategie des Managements, welche die Funktion (158) maximiert ( $\{c^{d*}(\theta), \hat{\theta}^{m*}(\theta)\}$  stellt gewissermassen den Pfad der optimalen Strategien entlang des Pfades der Fallschweren dar). Bei einem Patienten der Fallschwere  $\theta$  ist der Gewinn aus der optimalen Strategie  $\{c^{d*}(\theta), \hat{\theta}^{m*}(\theta)\}$  für das Management mindestens so hoch wie aus jeder anderen möglichen Strategie, sonst wäre sie ja nicht optimal. Insbesondere gilt dies im Vergleich mit der Strategie  $\{c^{d*}(\tilde{\theta}), \hat{\theta}^{m*}(\tilde{\theta})\}$ , die bei der Fallschwere  $\tilde{\theta} \neq \theta$  optimal wäre [vgl. Breyer (2005)]:

$$k(\hat{\theta}^{m*}(\theta)) - r(\hat{\theta}^{m*}(\theta))c^{d*}(\theta) + \theta V^*(c^{d*}(\theta)) - u \geq k(\hat{\theta}^{m*}(\tilde{\theta})) - r(\hat{\theta}^{m*}(\tilde{\theta}))c^{d*}(\tilde{\theta}) + \theta V^*(c^{d*}(\tilde{\theta})) - u. \quad (159)$$

Nun kann der Auftraggeber ein direktes Vergütungssystem einsetzen, das für jede Fallschwere eine Vergütung  $t^m(\theta)$  sowie eine Kostenvorgabe  $c^m(\theta)$  bestimmt. Wenn der Auftraggeber die Parameter dieses direkten Systems definiert, sodass

$$t^m(\theta) = k(\hat{\theta}^{m*}(\theta)) - r(\hat{\theta}^{m*}(\theta))c^{d*}(\theta) \text{ und } c^m(\theta) = c^{d*}(\theta), \quad (160)$$

gilt, dann ergibt Einsetzen in (159)

$$t^m(\theta) + \theta V^*(c^m(\theta)) \geq t^m(\tilde{\theta}) + \theta V^*(c^m(\tilde{\theta})). \quad (161)$$

Dieses direkte Vergütungssystem veranlasst das Management also, die Fallschwere dem Auftraggeber wahrheitsgemäss zu berichten und gewährt dem Management dieselbe Budgetierung, die es unter dem cost-sharing System von sich aus gewählt hätte, so dass auch dieselbe Allokation resultiert.

Wenn jede Allokation, die aus einem indirektem Vergütungssystem wie der linearen Kostenbeteiligung resultiert, auch durch ein direktes Vergütungssystem erreicht wird, dann bedeutet dies im Umkehrschluss, dass eine bestimmte Allokation überhaupt nur dann zu erreichen ist, wenn sie durch ein direktes Vergütungssystem implementiert werden kann. Dies gilt insbesondere für die bestmögliche Allokation.

Um die bestmögliche Allokation zu bestimmen, bedarf es also „lediglich“ der Bestimmung der optimalen Parameter  $t^m(\theta)$  und  $c^m(\theta)$  des direkten Vergütungssystems, unter der Bedingung (161), die sicherstellt, dass das Management die Fallschwere wahrheitsgemäss



berichtet (zudem muss natürlich auch noch die Teilnahmebedingung des Managements beachtet werden). Die Bedingung (161) stellt also die in Abschnitt 3 erläuterte Anreizverträglichkeitsbedingung (4) dar.

In der theoretischen Analyse der optimalen Spitalregulierung bei asymmetrischer Information bezüglich der Fallschwere findet dieser Ansatz zum Beispiel in den Beiträgen von Siciliani (2006) bzw. Chalkley und Malcomson (2002) Anwendung.

### 8.1.3 Asymmetrische Information hinsichtlich der Fallschwere zwischen Management und Auftraggeber im Prinzipal-Agent-Modell mit doppeltem Unterstellungsverhältnis

**8.1.3.1 Annahmen bezüglich der Informationsverteilung** Im Folgenden wollen wir untersuchen, wie sich die asymmetrische Informationsverteilung bezüglich der Fallschwere im PA-Modell mit doppeltem Unterstellungsverhältnis auswirkt. Zur Vereinfachung nehmen wir aber wie in Kapitel 6 an, die Anstrengungen zur Verbesserung der Kosteneffizienz seien exogen gegeben. In der folgenden Analyse gehen wir zunächst davon aus, dass der Arzt dem Management die Fallschwere immer wahrheitsgemäss berichtet und die Dokumentationen in der Patientenakte den Tatsachen entsprechend führt. Vom Management hingegen nehmen wir an, dass es diese Information manipuliert an den Auftraggeber weiterleitet, wenn dies von Vorteil für das Management ist, d.h. wenn es dadurch seinen Gewinn steigern kann. Tabelle 3 gibt einen Überblick über die Verteilung der Information.

**8.1.3.2 Direktes Vergütungssystem** Der Fokus der Analyse liegt primär auf der Frage, welche Allokation der Auftraggeber bestenfalls erreichen kann, wenn das Management die Fallschwere in seinem Bericht gegebenenfalls falsch darstellt. Mit dem Prinzip der direkten Offenbarung genügt es daher, die optimale Allokation zu bestimmen, die mit dem direkten Vergütungssystem möglich ist. Wir nehmen also an, der Auftraggeber setzt ein System ein, das mit  $\{t^m(\theta), c^m(\theta)\}$  pro Patient die Vergütung und die Budgetvorga-

<i>Grösse beobachtbar ja/nein</i>	Auftraggeber	Management	Arzt
Mengen $q_1, q_2$	nein	nein	ja
Preise $p_1, p_2$	ja	ja	ja
Kosten $c$	ja	ja	ja
Fallschwere $\theta$	nein	ja	ja

Tabelle 3: Informationsverteilung des Modells in Abschnitt 8.1.3.

be für das Management in Abhängigkeit der (vom Management berichteten) Fallschwere bestimmt.

An dieser Stelle ist anzumerken, dass mit dem direkten Vergütungssystem  $\{t^m(\theta), c^m(\theta)\}$  dem Management die Möglichkeit genommen wird, das interne Fallbudget  $c^d(\theta)$  selbst zu bestimmen. Vielmehr wird das interne Fallbudget mit der gemeldeten Fallschwere festgelegt. Wenn das Management für die Fallschwere  $\theta$  die Fallschwere  $\hat{\theta}^m$  meldet, muss es das interne Fallbudget bei  $c^d(\theta) = c^m(\hat{\theta}^m)$  setzen. Wenn das Management wahrheitsgemäss berichtet, muss also  $c^d(\theta) = c^m(\theta)$  gelten.

**8.1.3.3 Verteilung und Dichte der Fallschwere** Aus technischen Gründen nehmen wir im Zusammenhang mit der Verteilung der Fallschwere  $F(\theta)$  an, dass die sogenannte „Inverse Hazard Rate“  $h(\theta) = \frac{1-F(\theta)}{f(\theta)}$  streng monoton fällt:

$$h'(\theta) = \frac{d\left(\frac{1-F(\theta)}{f(\theta)}\right)}{d\theta} < 0. \quad (162)$$

Die Eigenschaft der „monotonen Hazard Rate“ wird von den meisten Verteilungen mit eingipfelter Dichte erfüllt [Laffont/Martimort (2002), Seite 137]. Die Annahme ist Standard in Modellen, die sich mit der Problematik der versteckten Information befassen [z.B. Laffont/Tirole (1993)].

**8.1.3.4 Optimale Allokation bei vollständiger Information über die Fallschwere** Um den Einfluss der asymmetrischen Information auf die erreichbare Allokation zu identifizieren, ist ein Vergleich mit der optimalen Allokation hilfreich, die der Auftraggeber erreichen kann, wenn er über die Fallschwere vollständig informiert ist, die Mengen für ihn aber nicht zu beobachten sind, wie es in Kapitel 6 angenommen wurde. Als Referenz-Allokation dient also jene aus Abschnitt 6.3. Die Allokation hinsichtlich der optimalen Budgetierung  $c^d(\theta)$  für die Behandlung eines Patienten der Fallschwere  $\theta$  ist durch

$$\left[ \theta B_{q_1}^*(.) \frac{\partial q_1^*(.)}{\partial c^d(.)} + \theta B_{q_2}^*(.) \frac{\partial q_2^*(.)}{\partial c^d(.)} \right] + \lambda^*(\theta, .) - 1 = 0. \quad (163)$$

beschrieben. Hinsichtlich des Nutzenüberschusses von Arzt und Management muss

$$N\ddot{U}(\theta) = g \quad \forall \theta \in \Theta \quad (164)$$

erfüllt sein.

### 8.1.4 Lösung des Modells bei spitalinterner Prinzipal-Agent-Lösung

#### 8.1.4.1 Zeitliche Abfolge

Der zeitliche Ablauf des Spiels ist wie folgt:

1. In der ersten Stufe des Spieles bestimmt der Auftraggeber die Parameter  $t^m(\theta)$  und  $c^m(\theta)$  für jede Fallschwere  $\theta \in \Theta$ .
2. In der zweiten Stufe entscheidet das Management über die Annahme oder die Ablehnung des Vertragsangebotes des Auftraggebers. Wenn es annimmt, bestimmt das Management für jede Fallschwere  $\theta \in \Theta$  seine Strategie der Weiterleitung der später vom Arzt berichteten Fallschwere  $\hat{\theta}^m(\theta)$  (und damit implizit die Fallbudgetierung  $c^d(\theta)$ ), sowie die freien Mittel für den Arzt  $t^d(\theta)$ .
3. In der dritten Stufe kommen die  $n$  Patienten zur Behandlung in das Spital. Für jeden einzelnen Patienten stellt der Arzt die Fallschwere  $\theta$  fest und entscheidet, ob er den Patienten zur Behandlung aufnimmt oder ob er dem Patienten die Aufnahme verweigert. Wenn er sich für die Aufnahme entscheidet, teilt der Arzt dem Management die Fallschwere mit. Das Management entscheidet, ob es die Behandlung des konkreten Patienten zulässt. Wenn das Management die Behandlung zulässt, wählt der Arzt die Mengen  $q_1$  und  $q_2$  für die Behandlung des Patienten so, dass das vom Management vorgegebene Budget  $c^d(\theta)$  eingehalten wird. Das Management berichtet die Fallschwere gemäss seiner in Stufe 2 gewählten Strategie an den Auftraggeber. Die Behandlungskosten werden von allen drei Akteuren beobachtet. Die Zahlungen finden statt.

Wir lösen das Modell mittels Rückwärtsinduktion.

**8.1.4.2 Bestimmung der Mengen  $q_1$  und  $q_2$  durch den Arzt** Für den Arzt ändert sich im Vergleich zu Abschnitt 6.3.2 nichts: Bei der Behandlung eines Patienten der Fallschwere  $\theta$  maximiert er seinen Nutzen mit der Wahl der Mengen  $q_1$  bzw.  $q_2$  bei gegebenem Budget  $c^d(\theta)$ . Somit wird auch auf die Ergebnisse von Abschnitt 6.3.2 verwiesen.

Da die optimalen Mengen  $q_1^*(.)$  und  $q_2^*(.)$  Funktionen des Budgets sind, verwenden wir im Folgenden die vereinfachende Schreibweise

$$V^*(c^d(\theta)) \equiv V(q_1^*(.), q_2^*(.)) \quad (165)$$

wenn die Abhängigkeit vom Budget betont werden soll.

**8.1.4.3 Anreizbedingung für wahrheitsgemässe Berichterstattung durch das Management** Das Management muss entscheiden, wie es den Bericht des Arztes über die Fallschwere an den Auftraggeber weiterleitet. Sei  $\hat{\theta}^m$  der Bericht des Managements an den Auftraggeber. Für den Gewinn des Managements bei der Behandlung eines Patienten der Fallschwere  $\theta$  ergibt sich damit

$$G(\theta, \hat{\theta}^m) = t^m(\hat{\theta}^m) - t^d(\theta). \quad (166)$$

Wie bereits weiter oben vermerkt, legt sich das Management mit seinem Bericht an den Auftraggeber auf die Fallbudgetierung fest, die der Auftraggeber für die gemeldete Fallschwere vorsieht, d.h.  $c^d(\theta) = c^m(\hat{\theta}^m)$ . Der Gewinn des Managements (166) lässt sich mit der Nutzenfunktion des Arztes äquivalent schreiben als

$$G(\theta, \hat{\theta}^m) = t^m(\hat{\theta}^m) + \theta V^*(c^m(\hat{\theta}^m)) - U(\theta). \quad (167)$$

Da der Nutzen des Arztes negativ in den Gewinn eingeht, gesteht das Management dem Arzt gerade dessen Reservationsnutzen zu. Mit  $U(\theta) = u$  ergibt sich für das Maximierungsproblem des Managements für einen Patienten der Fallschwere  $\theta$

$$\max_{\hat{\theta}^m} G(\theta, \hat{\theta}^m) = t^m(\hat{\theta}^m) + \theta V^*(c^m(\hat{\theta}^m)) - u. \quad (168)$$

Sei  $G(\theta) \equiv G(\theta, \theta)$  der Gewinn des Managements bei wahrheitsgemässer Berichterstattung. Wenn wahrheitsgemässe Berichterstattung für das Management optimal ist, d.h.  $\hat{\theta}^{m*}(\theta) = \theta$ , dann ergibt sich mit dem Envelope-Theorem:

$$\dot{G}(\theta) \equiv \frac{dG(\theta)}{d\theta} = V^*(c^m(\theta)) > 0. \quad (169)$$

Gleichung (169) besagt, wenn wahrheitsgemässe Berichterstattung bei einer Fallschwere  $\theta$  optimal ist, dann steigt der Gewinn mit der Fallschwere im Bereich um  $\theta$ . Die Notation  $\dot{G}(\theta) \equiv \frac{dG(\theta)}{d\theta}$  betont, dass es sich hier um einen Pfad von  $G(\theta)$  entlang der Fallschweren handelt. Diese Notation wird im Folgenden für alle Variablen entsprechend verwendet.

Gleichung (169) stellt die Anreizbedingung für eine wahrheitsgemässe Berichterstattung durch das Management dar.<sup>67</sup> Sie verlangt, dass der vom Auftraggeber eingesetzte Mechanismus  $\{t^m(\theta), c^m(\theta)\}$  dem Management bei der Behandlung eines Patienten mit der Fallschwere  $\theta$  einen Gewinn ermöglicht, der um  $V^*(c^d(\theta))d\theta$  Einheiten über jenem Gewinn

---

<sup>67</sup>(169) ist eine äquivalente Darstellung der Anreizbedingung (161).

liegt, den das Management bei einem Patienten der Fallschwere  $\theta' = \theta - d\theta$  erreicht (mit  $d\theta > 0$ ).

In Anhang (A.4) wird gezeigt, dass die Bedingung (169) zusammen mit

$$\dot{c}^m(\theta) \geq 0 \quad \forall \theta \in \Theta \quad (170)$$

ein globales Maximum identifiziert. Wenn (169) und (170) für alle Fallschweren  $\theta \in \Theta$  erfüllt sind, dann ist wahrheitsgemässe Berichterstattung immer die optimale Strategie des Managements. Wir gehen zunächst davon aus, dass die Bedingung (170) erfüllt ist und überprüfen weiter unten, ob dies tatsächlich der Fall ist, wenn der Auftraggeber das optimale System wählt.

Die Intuition hinter Gleichung (169) ist folgende: Angenommen der Arzt behandelt zwei Patienten, der erste Patient hat die Fallschwere  $\theta$ , der zweite hat die marginal höhere Fallschwere  $\theta + d\theta$  mit  $d\theta > 0$ . Nehmen wir an, das Management berichte nur für den Patienten der höheren Fallschwere wahrheitsgemäss, für den Patienten mit der geringeren Fallschwere gibt es aber an, dieser sei ebenfalls von der Fallschwere  $\theta + d\theta$ . Dann erhält das Management zwar für beide Fälle das gleiche Budget  $c^m(\theta + d\theta)$  und die gleiche Vergütung  $t^m(\theta + d\theta)$ , muss aber beim Patienten der tieferen Fallschwere dem Arzt, damit dieser seinen Reservationsnutzen  $u$  erhält, freie Mittel zur Verfügung stellen, die um den Betrag  $d\theta V^*(c^m(\theta + d\theta))$  höher sein müssen als beim Patienten der marginal höheren Fallschwere. Der Gewinn des Managements beim Patienten mit der Fallschwere  $\theta$  ist also um den Betrag  $d\theta V^*(c^m(\theta + d\theta))$  tiefer als beim Patienten mit der Fallschwere  $\theta + d\theta$ :

$$G(\theta + d\theta, \theta + d\theta) - G(\theta, \theta + d\theta) = d\theta V^*(c^m(\theta + d\theta)). \quad (171)$$

Damit das Management auch für den Patienten mit der Fallschwere  $\theta$  wahrheitsgemäss berichtet, muss diese Option mindestens den gleichen Gewinn erbringen wie der zu hohe Bericht  $\hat{\theta}^m = \theta + d\theta$ . Aus Gleichung (171) folgt, dass das Management gerade indifferent zwischen dem zu hohen Bericht  $\hat{\theta}^m = \theta + d\theta$  und dem wahrheitsgemässen Bericht  $\hat{\theta}^m = \theta$  ist, wenn der Gewinn bei wahrheitsgemäßer Berichterstattung beim Patienten der Fallschwere  $\theta$  genau um den Betrag  $d\theta V^*(c^m(\theta + d\theta))$  tiefer ist als beim Patienten mit der marginal höheren Fallschwere  $\theta + d\theta$ . Bedingung (169) folgt aus (171) mit  $d\theta \rightarrow 0$ :

$$\lim_{d\theta \rightarrow 0} \frac{G(\theta + d\theta, \theta + d\theta) - G(\theta, \theta + d\theta)}{d\theta} = \dot{G}(\theta) = V^*(c^m(\theta)). \quad (172)$$

#### 8.1.4.4 Maximierungsproblem des Auftraggebers und optimale Allokation

Der ex-post Nutzen des Auftraggebers bei der Behandlung eines Patienten der Fallschwere ist gegeben mit

$$\theta B^*(c^m(\theta)) - c^m(\theta) - t^m(\theta) = \theta B^*(c^m(\theta)) + \theta V^*(c^m(\theta)) - c^m(\theta) - N\ddot{U}(\theta) - u \quad (173)$$

wobei  $B^*(c^m(\theta)) \equiv B(q_1^*(.), q_2^*(.))$ . Der Auftraggeber muss das Vergütungssystem  $\{t^m(\theta), c^m(\theta)\}$  allerdings festlegen, bevor die Fallschwere bekannt wird. Daher ist er an der Maximierung des erwarteten Nutzens interessiert

$$\max_{c^m(\theta), N\ddot{U}(\theta)} E[W(\theta)] = \int_{\underline{\theta}}^{\bar{\theta}} [\theta B^*(c^m(\theta)) + \theta V^*(c^m(\theta)) - c^m(\theta) - N\ddot{U}(\theta)] f\theta d\theta - u. \quad (174)$$

Dabei muss er allerdings beachten, dass der Vertrag die mit Gleichung (169) formulierte Anreizkompatibilität sowie die Teilnahmebedingung  $N\ddot{U}(\theta) \geq g \quad \forall \theta$  des Managements erfüllt.

Bei spitalinternem PA-Verhältnis gewährt das Management dem Arzt immer gerade seinen Reservationsnutzen, sodass der gemeinsame Nutzenüberschuss dem Gewinn des Managements entspricht:  $N\ddot{U}(\theta) = G(\theta)$ . Da wegen der Anreizbedingung (169) der Gewinn des Managements (bzw. der Nutzenüberschuss) in der Fallschwere zunimmt, gilt für alle  $\theta > \underline{\theta}$ ,  $G(\theta) \geq G(\underline{\theta})$  bzw.  $N\ddot{U}(\theta) \geq N\ddot{U}(\underline{\theta})$ . Daher impliziert die Teilnahme des Managements bei einem Patienten mit der niedrigsten Fallschwere  $\theta = \underline{\theta}$  die Teilnahme bei allen anderen Fallschweren. Und da der Auftraggeber den Nutzenüberschuss immer so niedrig als möglich halten möchte, gewährt er dem Management bei einem Patienten der niedrigsten Fallschwere gerade den kleinstmöglichen Überschuss

$$N\ddot{U}(\underline{\theta}) = g. \quad (175)$$

Der Auftraggeber maximiert also seinen erwarteten Nutzen (174) unter Berücksichtigung der Anreizkompatibilitätsbedingung (169) sowie der Teilnahmebedingung (175). Um die Maximierung etwas übersichtlicher zu gestalten, werden die beiden Bedingungen (169) und (175) zusammengefasst. Für den Nutzenüberschuss bei der Behandlung eines Patienten der Fallschwere  $\theta$  ergibt sich mit den beiden Bedingungen

$$N\ddot{U}(\theta) = \int_{\underline{\theta}}^{\theta} V^*(c^m(\tilde{\theta})) d\tilde{\theta} + g. \quad (176)$$

Für den erwarteten Gewinn bzw. Nutzenüberschuss erhalten wir

$$E[N\ddot{U}(\theta)] = \int_{\underline{\theta}}^{\bar{\theta}} \int_{\underline{\theta}}^{\theta} V^*(c^m(\tilde{\theta})) d\tilde{\theta} f(\theta) d\theta + g. \quad (177)$$

Partielle Integration führt zu

$$\begin{aligned} E[N\ddot{U}(\theta)] &= [F(\theta) \int_{\underline{\theta}}^{\theta} V^*(c^m(\tilde{\theta})) d\tilde{\theta}]|_{\underline{\theta}}^{\bar{\theta}} - \int_{\underline{\theta}}^{\bar{\theta}} V^*(c^m(\theta)) F(\theta) d\theta + g \\ &= \int_{\underline{\theta}}^{\bar{\theta}} V^*(c^m(\theta)) d\theta - \int_{\underline{\theta}}^{\bar{\theta}} V^*(c^m(\theta)) F(\theta) d\theta + g \\ &= \int_{\underline{\theta}}^{\bar{\theta}} h(\theta) V^*(c^m(\theta)) f(\theta) d\theta + g \end{aligned} \quad (178)$$

mit  $h(\theta) \equiv \frac{1-F(\theta)}{f(\theta)}$ , der inversen Hazard-Rate. Der erwartete Nutzenüberschuss (178) lässt sich nun einfach in das Maximierungsproblem (174) des Auftraggebers einsetzen. Wir erhalten dadurch:

$$\max_{c^m(\theta)} E[W(\theta)] = \int_{\underline{\theta}}^{\bar{\theta}} [\theta(B^*(c^m(\theta)) + V^*(c^m(\theta))) - c^m(\theta) - h(\theta)V^*(c^m(\theta))] f(\theta) d(\theta) - g - u. \quad (179)$$

Wir nehmen an, der Integrand von (179) sei konkav und stetig für alle Fallschweren  $\theta$ . Dann entspricht die Maximierung von

$$\max_{c^m(\theta)} \tilde{W}(\theta) = \theta(B^*(c^m(\theta)) + V^*(c^m(\theta))) - c^m(\theta) - h(\theta)V^*(c^m(\theta)). \quad (180)$$

für jede Fallschwere der Maximierung von (179) (Punktweise Maximierung). Gleichung (180) lässt sich als Maximierungsproblem eines Auftraggebers interpretieren, der die Fallschwere beobachten und vertraglich verwerten kann, für den sich aber die Kosten der Beobachtung der Fallschwere auf  $h(\theta)V^*(c^m(\theta))$  belaufen [vgl. Laffont/Martimort (2002), Seite 48].

Als Bedingung erster Ordnung für das optimale Fallbudget  $c^{m*}(\theta)$  bei einer Fallschwere von  $\theta$  erhalten wir (mit  $\frac{\theta}{\theta} \frac{\partial V^*(c^m(\theta))}{\partial c^m} = \frac{\lambda^*(\theta, \cdot)}{\theta}$ )

$$\left( \theta B_{q_1}^* \frac{\partial q_1^*(\cdot)}{\partial c(\theta)} + \theta B_{q_2}^* \frac{\partial q_2^*(\cdot)}{\partial c(\theta)} \right) + \left( 1 - \frac{h(\theta)}{\theta} \right) \lambda^*(\theta, \cdot) = 1. \quad (181)$$

Für die optimale Vergütung folgt mit (176)

$$t^m(\theta) = \int_{\underline{\theta}}^{\theta} V^*(c^m(\tilde{\theta}))d\tilde{\theta} - \theta V^*(c^{m*}(\theta)) + g. \quad (182)$$

Zusammen mit der Anreizkompatibilitätsbedingung (169) bzw. des Nutzenüberschusses (176), die für jede Fallschwere den Nutzenüberschuss  $N\ddot{U}(\theta)$  bzw. den Gewinn  $G(\theta)$  des Managements definiert, beschreibt Gleichung (181) das Standard-Ergebnis des PA-Modells mit der Problematik versteckter Information: Der Prinzipal muss einen Trade-off zwischen einer effizienten Allokation der endogenen Aktionsparameter und der Informationsrente des Agenten eingehen, hier also zwischen effizienter Budgetierung und minimalem Nutzenüberschuss.

Lediglich bei einem Patienten der höchsten Fallschwere entspricht Bedingung (181) der Bedingung für die optimale Budgetierung (163), die ohne asymmetrische Information bezüglich der Fallschwere resultiert, denn bei  $\bar{\theta}$  ist der Ausdruck  $h(\bar{\theta})$  null. Allerdings ist die Informationsrente wegen der Anreizbedingung (169) bei  $\theta = \bar{\theta}$  maximal. Der Trade-off geht bei einem Patienten der höchsten Fallschwere also vollständig auf die Informationsrente, während die Budgetierung effizient ist. Bei einem Patienten der niedrigsten Fallschwere gilt das Gegenteil. Zwar erhält das Management dann keine Informationsrente, allerdings ist die Budgetierung nach unten verzerrt. Bei allen Fallschweren zwischen der höchsten und der niedrigsten äussert sich der Trade-off in einer Verzerrung der Budgetierung nach unten und in einer Informationsrente, die umso höher ist, je höher die Fallschwere ist.

Es fehlt noch der Nachweis, dass die Bedingung zweiter Ordnung  $\dot{c}^{m*}(\theta) \geq 0$  (170) erfüllt ist. Da wir annehmen, (180) sei streng konkav, muss die zweite Ableitung von (180) negativ sein. Implizites Differenzieren von 181 ergibt

$$\frac{dc}{d\theta} = - \frac{\left[ B_{q_1}^*(.) \frac{\partial q_1^*(.)}{\partial c^d(.)} + B_{q_2}^*(.) \frac{\partial q_2^*(.)}{\partial c^d(.)} \right] + \left( 1 - \frac{h(\theta)}{\theta} \right) \frac{\partial \lambda^*(\theta, .)}{\partial \theta} - \frac{d(\frac{h(\theta)}{\theta})}{d\theta} \lambda^*(\theta, .)}{\text{Zweite Ableitung}}. \quad (183)$$

Die Bedingung  $\dot{c}^{m*}(\theta) \geq 0$  ist also erfüllt, wenn

$$\left[ B_{q_1}^*(.) \frac{\partial q_1^*(.)}{\partial c^d(.)} + B_{q_2}^*(.) \frac{\partial q_2^*(.)}{\partial c^d(.)} \right] + \left( 1 - \frac{h(\theta)}{\theta} \right) \frac{\partial \lambda^*(\theta, .)}{\partial \theta} - \frac{d(\frac{h(\theta)}{\theta})}{d\theta} \lambda^*(\theta, .) \geq 0 \quad (184)$$

für alle Fallschweren gilt. Da  $\frac{d(\frac{h(\theta)}{\theta})}{d\theta} = \frac{h'(\theta)\theta - h(\theta)}{\theta^2} < 0$ , ist dies dann eindeutig gewährleistet, wenn  $(1 - \frac{h(\theta)}{\theta}) \geq 0 \forall \theta \in \Theta$  der Fall ist. Wenn dagegen für einige Fallschweren  $(1 - \frac{h(\theta)}{\theta}) < 0$  gilt, was bei niedrigeren Fallschweren wahrscheinlicher ist als bei hohen, dann bindet womöglich die Bedingung  $\dot{c}^{m*}(\theta) \geq 0$ . Ist dies der Fall, resultiert für ein oder



mehrere Intervall/e der Fallschwere eine konstante Budgetierung („Bunching“, vgl. Laffont/Martimort [Laffont/Martimort (2002), Seite 140]). Wir nehmen hier allerdings an, dass  $\dot{c}^{m*}(\theta) > 0$  für alle Fallschweren erfüllt ist und fassen die Ergebnisse zusammen:

**Ergebnis 8.1** *Die Möglichkeit des Managements, die Fallschwere gegebenenfalls falsch zu berichten, zwingt den Auftraggeber zu einem Trade-off zwischen Zugeständnissen bei dem Nutzenüberschuss, den er dem Management und dem Arzt gewährt, und der optimalen Fallbudgetierung. Für Patienten einer Fallschwere unterhalb der höchsten resultiert eine Verzerrung der Budgetierung nach unten. Gleichzeitig erhält das Management einen Nutzenüberschuss, der bei Patienten mit einer Fallschwere oberhalb der niedrigsten grösser ist als der Reservationswert.*

## 8.2 Spitalinterne Informationsasymmetrie hinsichtlich der Fallschwere

Wir haben gesehen, dass es aus Sicht des Auftraggebers zu Verzerrungen bei der Budgetierung und des Nutzenüberschusses bzw. dem Gewinn des Managements kommt, wenn das Verhältnis zwischen Auftraggeber und Management von asymmetrischer Information bezüglich der Fallschwere geprägt ist und das Management seinen Informationsvorsprung zu seinen Gunsten ausnutzt.

Aber wie ist es um die internen Verhältnisse zwischen Management und Arzt bestellt? Das Management (in Form der Kodierabteilung) wertet das Patientendossier aus und bestimmt die entsprechenden ICD-10- bzw. CHOP-Kodes. Was jedoch im Dossier dokumentiert ist, dafür ist letzten Endes doch vor allem der Arzt verantwortlich. Selbstverständlich existieren bei der Dokumentation Spielräume für den Arzt, deren Ausschöpfung vom Management nicht zu beobachten ist, so dass grundsätzlich auch zwischen Arzt und Management eine Informationsasymmetrie bezüglich der Fallschwere zu vermuten ist.

Wenn nun die Fallschwere, die sich aus der vom Arzt verfassten Dokumentation ergibt, den Nutzen des Arztes in irgendeiner Weise beeinflusst (zum Beispiel wenn die Budgetierung seiner Abteilung von der Fallschwere abhängt), warum sollte der Arzt dann seinen Informationsvorsprung nicht ausnutzen und die Diagnosen und Behandlungen des Patienten im Dossier so darstellen, dass die daraus von der Kodierabteilung abzuleitende Fallschwere zu seinem Vorteil ist? Natürlich soll hier der Ärzteschaft nicht die Neigung zu Betrug unterstellt werden. Es geht viel mehr wieder um das Ausschöpfen von Spielräumen, die auch Raum für ökonomische Beweggründe lässt. Oder wie Mark Pauly, Professor für Health Care Management in Pennsylvania, formuliert: „*Other things equal, physicians would rather tell the truth, but they would be willing to surrender some accuracy for so-*

*me amount of money income*“ [Pauly zitiert in McGuire (2000), Seite 507]. Verfolgt man diesen Gedanken konsequent weiter, so erscheint es zumindest möglich, dass durch die interne Informationsasymmetrie bezüglich der Fallschwere eine zweite Quelle für Verzerrungen entsteht. So halten wir fest:

**Folgerung 8.2** *Auch zwischen Arzt und Management ist asymmetrische Information bezüglich der Fallschwere zu vermuten.*

Wenn der Arzt sich durch das Ausnutzen von Spielräumen einen Vorteil verschaffen kann, dann kann dies der Grund für Verzerrungen bei der spitalinternen Allokation sein, die auch für den Auftraggeber relevante Konsequenzen haben können. In Abschnitt 6 haben wir gesehen, dass die asymmetrische Information zwischen Arzt und Management hinsichtlich der vom Arzt zur Behandlung eingesetzten Mengen  $q_1$  und  $q_2$  für den Auftraggeber ohne Bedeutung ist, solange das Management keine eigenen Präferenzen gegenüber der Behandlungsintensität besitzt. Wir stellen uns nun die Frage, ob es dagegen für den Auftraggeber eine Rolle spielt, wenn es spitalintern zwischen Management und Arzt zu Informationsasymmetrien bezüglich der Fallschwere kommt. Ebenso wie der Auftraggeber beim Management muss das Management nun seinerseits eine Anreizbedingung des Arztes beachten. Ist es nun von Relevanz für den Auftraggeber, wenn das Management die wahrheitsgemässe Berichterstattung vom Arzt „erkaufen“ muss, in dem es dafür sorgt, dass die Anreizbedingung des Arztes erfüllt ist, um darauf die Information an den Auftraggeber zu „verkaufen“? Oder, wie es McAfee und McMillan formulieren: „Does it matter how many hands information passes through between information source and decision maker?“ [McAfee/McMillan (1995)].

In der Literatur über die optimale Vergütung von Spitalleistungen finden sich keine Arbeiten über diese Problematik. In der allgemeinen Industrieökonomik hat der Einfluss dieses „bottom-up“ Informationsflusses über mehrere Hierarchieebenen dagegen einige Aufmerksamkeit erhalten [Mookherjee (2006)].

Das Grundproblem bei der Weiterleitung von Informationen über exogene Grössen über mehrere Hierarchieebenen wird von McAfee/McMillan (1995) beschrieben: Die Autoren zeigen, dass die durch die asymmetrische Information bedingte Verzerrung mit der Anzahl der Hierarchiestufen, über welche die Information geleitet wird, zunimmt. In ihrem hierarchischen PA-Modell bilden die Autoren eine Situation ab, in welcher der Agent auf unterster Hierarchieebene für die Produktion eines Outputs verantwortlich ist. Für die Steuerung des Unternehmens ist der Prinzipal auf der obersten Hierarchiestufe verantwortlich. Der Prinzipal hat das Problem, dass nur der Agent über produktionskostenrelevante Faktoren informiert ist und die Kommunikation zwischen Prinzipal und Agent nur indirekt über eine Reihe von Intermediären möglich ist, die hierarchisch zwischen Agent

und Prinzipal organisiert sind. Lediglich der Intermediär, der dem Agent direkt überstellt ist, kommuniziert mit dem Agenten und definiert dessen Vergütung. Dieser unterstellte Intermediär ist seinerseits dem Intermediär auf der nächsten Hierarchiestufe unterstellt und wird von diesem vergütet usw.. Möchte am unteren Ende der Hierarchie der unterste Intermediär an wahrheitsgemässe Informationen über die Produktionsfaktoren gelangen, muss er dafür dem Agenten eine Informationsrente zugestehen. Mit dem wahrheitsgemässen Bericht ist der unterste Intermediär aber gegenüber seinem direkten Vorgesetzten Träger privater Information, was ihm seinerseits eine Informationsrente ermöglicht. So steigt nun die gesamte Informationsrente mit der Anzahl der Hierarchiestufen bis der Prinzipal an die Informationen gelangt. Dies führt aus Sicht des Prinzipals zu steigenden marginalen Gesamtkosten, bestehend aus den marginalen Produktionskosten sowie den marginalen Kosten der Informationsrenten, und damit zu einer steigenden unerwünschten Verzerrung des optimalen Outputs nach unten.

Wir wollen im Folgenden feststellen, ob es sich in unserem Modell auswirkt, wenn der Arzt die exogen gegebene Fallschwere intern an das Management falsch darstellen kann und es für das Management eine Anreizbedingung des Arztes zu beachten gilt. Sind die internen Verhältnisse der Informationsverteilung bezüglich der Fallschwere für den Auftraggeber von Relevanz? Wenn ja, wie wirkt sie sich aus?

Wir gehen dieser Frage nach, indem wir das Modell aus dem vorigen Abschnitt erneut lösen, sehen jedoch von der Annahme ab, dass der Arzt die Fallschwere immer wahrheitsgemäss an das Management berichtet.

### **8.2.1 Spitalinterne Informationsasymmetrie hinsichtlich der Fallschwere zwischen Arzt und Management im Prinzipal-Agent-Modell mit doppeltem Unterstellungsverhältnis**

**8.2.1.1 Informationsverteilung** Wir nehmen an, dass der Arzt die Fallschwere nur dann wahrheitsgemäss an das Management berichtet, wenn dies die optimale für den Arzt ist. Ansonsten gilt dieselbe Informationsverteilung wie in Abschnitt 8.1.3. Tabelle 4 gibt einen Überblick.

### **8.2.2 Lösung des Modells bei spitalinterner Prinzipal-Agent-Lösung**

Der Zeitliche Ablauf entspricht jenem aus Abschnitt 8.2.1. Abermals gehen wir die Analyse mittels Rückwärtsinduktion an.

**8.2.2.1 Entscheidungen des Arztes: Bestimmung der Mengen  $q_1$  und  $q_2$  und Strategie der Berichterstattung** Der Arzt entscheidet nun zum einen über die Men-

<i>Grösse beobachtbar ja/nein</i>	Auftraggeber	Management	Arzt
Mengen $q_1, q_2$	nein	nein	ja
Preise $p_1, p_2$	ja	ja	ja
Kosten $c$	ja	ja	ja
Fallschwere $\theta$	nein	nein	ja

Tabelle 4: Informationsverteilung des Modells in Abschnitt 8.2.1.

gen  $q_1$  und  $q_2$  und zum anderen darüber, welche Fallschwere er dem Management berichtet. Im Sinne der spieltheoretischen Modellierung fällt er die Entscheidungen in derselben Spielstufe. Zur mathematischen Analyse zerlegen wir die Entscheidungsfindung jedoch gedanklich in zwei Stufen.

In der ersten Stufe trifft der Arzt seine Entscheidungen bezüglich der zur Behandlung eingesetzten Mengen  $q_1$  und  $q_2$ . Dabei muss er wie gehabt die vom Management vorgegebene Budgetierung beachten. Insofern folgen dieselben Ergebnisse wie in Abschnitt 6.3.2. Allerdings sind die optimalen Mengen Funktionen von dem aus der gemeldeten Fallschwere  $\hat{\theta}^d$  resultierenden Budget  $c^d(\hat{\theta}^d)$ :

$$q_1^* \equiv q_1^*(c^d(\hat{\theta}^d)) \quad (185)$$

und

$$q_2^* \equiv q_2^*(c^d(\hat{\theta}^d)). \quad (186)$$

Im Unterschied zur bisherigen Analyse kann der Arzt nun die Budgetierung mit seinem Bericht beeinflussen und steht in der zweiten Stufe der Maximierung vor der Frage, welche Fallschwere  $\hat{\theta}^d$  er dem Management melden soll. Dabei wählt er bei einem Patienten der Fallschwere  $\theta$  den optimalen Bericht  $\hat{\theta}^{d*}$  so, dass sein Nutzen gegeben der Mengen  $q_1^*(c^d(\hat{\theta}^d))$  und  $q_2^*(c^d(\hat{\theta}^d))$  maximal wird. Die zweite Stufe seines Maximierungsproblems lautet also

$$\max_{\hat{\theta}^d} U(\theta, \hat{\theta}^d) = t^d(\hat{\theta}^d) + \theta V^*(c^d(\hat{\theta}^d)) \quad (187)$$

wobei  $\theta V^*(c^d(\hat{\theta}^d)) \equiv \theta V(q_1^*(c^d(\hat{\theta}^d)), q_2^*(c^d(\hat{\theta}^d)))$ .

Sei  $U(\theta) \equiv U(\theta, \theta)$  der Nutzen des Arztes, wenn er wahrheitsgemäss Bericht erstattet. Wenn wahrheitsgemässe Berichterstattung für den Arzt optimal ist, folgt mit dem Envelope-Theorem, dass

$$\dot{U}(\theta) = V^*(c^d(\theta)) > 0 \quad (188)$$

gelten muss. Die Bedingung (188) identifiziert zusammen mit

$$\dot{c}^d(\theta) \geq 0 \quad (189)$$

ein globales Maximum, d.h. der Arzt berichtet für alle Fallschweren wahrheitsgemäss (der Beweis erfolgt analog zu Anhang A.4). Wir gehen zunächst davon aus, dass die Bedingung (189) erfüllt ist.

**8.2.2.2 Teilnahmebedingung des Arztes** Bezüglich der Teilnahme des Arztes nehmen wir an, dass der Nutzen des Arztes mindestens den Reservationsnutzen  $u$  erreichen muss, ansonsten weist der Arzt die Patienten der entsprechende Fallschwere ab für die dies nicht der Fall ist. Damit alle Patienten behandelt werden, muss also

$$U(\theta) \geq u \quad \forall \theta \in \Theta \quad (190)$$

erfüllt sein. Da wegen der Anreizbedingung (188) der Nutzen des Arztes in der Fallschwere zunimmt, impliziert die Bereitschaft des Arztes zur Behandlung eines Patienten mit der niedrigsten Fallschwere  $\theta = \underline{\theta}$  die Teilnahme des Arztes bei allen anderen Fallschweren. Daher lässt sich die Teilnahmebedingung auch schreiben als

$$U(\underline{\theta}) \geq u. \quad (191)$$

**8.2.2.3 Entscheidungen des Managements: Berichterstattung an den Auftraggeber und Bestimmung der freien Mittel  $t^d(\theta)$**  Das Management bestimmt für jede Fallschwere zum einen die Berichtsfunktion  $\hat{\theta}^m(\theta)$ , d.h. es legt fest, ob es die vom Arzt gemeldete Fallschwere dem Auftraggeber wahrheitsgemäss weiterleiten oder dabei über- bzw. untertreiben soll. Zum anderen muss es die optimalen freien Mittel für den Arzt bestimmen. Bei der Budgetierung muss sich das Management dagegen nach der Budgetvorgabe des Auftraggebers richten, so dass  $c^d(\theta) = c^m(\hat{\theta}^m(\theta))$  gilt.

Der zeitliche Ablauf zwingt das Management jedoch, sich bezüglich der Meldestrategie sowie der freien Mittel festzulegen, bevor die Patienten zur Behandlung im Spital erscheinen und das Management die konkreten Fallschweren vom Arzt erfährt. Die Entscheidung zu

diesem Zeitpunkt bedeutet, dass das Management seinen erwarteten Gewinn maximiert. Dabei muss es beachten, dass der Arzt zur wahrheitsgemässen Berichterstattung veranlasst wird und am Vertrag überhaupt teilnimmt. Wahrheitsgemässe Berichterstattung des Arztes erreicht das Management, indem es die Anreizkompatibilitätsbedingung (188) bei der Maximierung ins Kalkül zieht, für die Teilnahme muss Bedingung (191) erfüllt sein. Das Maximierungsproblem des Managements in seiner grundsätzlichen Form stellt sich also wie folgt dar

$$\begin{aligned} \max_{\hat{\theta}^m(\theta), t^d(\theta)} E[G(\theta)] &= \int_{\underline{\theta}}^{\bar{\theta}} [t^m(\hat{\theta}^m(\theta)) - t^d(\theta)] f(\theta) d\theta. \\ s.t. \\ \dot{U}(\theta) &= V^*(c^m(\hat{\theta}^m(\theta))), \quad U(\underline{\theta}) \geq u. \end{aligned} \quad (192)$$

Das Maximierungsproblem (192) des Managements lässt sich analog zu (167) auch äquivalent formulieren:

$$\begin{aligned} \max_{\hat{\theta}^m(\theta), U(\theta)} E[G(\theta)] &= \int_{\underline{\theta}}^{\bar{\theta}} [t^m(\hat{\theta}^m(\theta)) + \theta V^*(c^m(\hat{\theta}^m(\theta)))] f(\theta) d\theta - \int_{\underline{\theta}}^{\bar{\theta}} U(\theta) f(\theta) d\theta \\ s.t. \\ \dot{U}(\theta) &= V^*(c^m(\hat{\theta}^m(\theta))), \quad U(\underline{\theta}) \geq u. \end{aligned} \quad (193)$$

Die Anreizbedingung (188) und die (191) lassen sich zusammenfassen und in die Zielfunktion einsetzen. Wenn die Teilnahmebedingung (191) bindet, lässt sich für jede Fallschwere der Nutzen des Arztes bestimmen, der resultiert, wenn Anreizbedingung (188) erfüllt ist:

$$U(\theta) = \int_{\underline{\theta}}^{\theta} V^*(c^m(\hat{\theta}^m(\tilde{\theta}))) d\tilde{\theta} + u. \quad (194)$$

Für den Erwartungsnutzen des Arztes ergibt sich damit

$$E[U(\theta)] = \int_{\underline{\theta}}^{\bar{\theta}} \int_{\underline{\theta}}^{\theta} V^*(c^m(\hat{\theta}^m(\tilde{\theta}))) d\tilde{\theta} f(\theta) d\theta + u. \quad (195)$$

Partielle Integration von (195) ergibt (vgl. (178))

$$E[U(\theta)] = \int_{\underline{\theta}}^{\bar{\theta}} h(\theta) V^*(c^m(\hat{\theta}^m(\theta))) f(\theta) d\theta + u \quad (196)$$

mit  $h(\theta) \equiv \frac{1-F(\theta)}{f(\theta)}$ . Mit Einsetzen von (196) in (193) lässt sich das Optimierungsproblem des Managements formulieren als

$$\max_{\hat{\theta}^m(\theta)} E[G(\theta)] = \int_{\underline{\theta}}^{\bar{\theta}} [t^m(\hat{\theta}^m(\theta)) + \theta V^*(c^m(\hat{\theta}^m(\theta))) - h(\theta)V^*(c^m(\hat{\theta}^m(\theta)))] f(\theta) d\theta - u. \quad (197)$$

Wir nehmen an, der Integrand von (197) sei konkav und stetig für alle Fallschweren  $\theta$ . Dann entspricht das Maximierungsproblem

$$\max_{\hat{\theta}^m(\theta)} \tilde{G}(\theta) = t^m(\hat{\theta}^m) + \theta V^*(c^m(\hat{\theta}^m)) - h(\theta)V^*(c^m(\hat{\theta}^m)) \quad (198)$$

für jede Fallschwere dem Maximierungsproblem (193) bzw. (197) (Punktweise Maximierung). Das Maximierungsproblem des Managements lässt sich also als ein Maximierungsproblem interpretieren, bei dem das Management die Fallschwere beobachten kann, dem Arzt aber einen Nutzenzuschlag in Höhe  $h(\theta)V^*(c^m(\hat{\theta}^m))$  über dem Reservationsnutzen  $u$  gewähren müsste. In diesem Sinne lässt sich (198) auch als „virtueller“ Gewinn des Managements interpretieren mit  $h(\theta)V^*(c^m(\theta))$  als „virtuelle“ Kosten, die dem Management durch die Beschaffung des wahrheitsgemässen Berichtes des Arztes über die Fallschwere entstehen [vgl. McAfee/McMillan (1995)].

**8.2.2.4 Bedingung für wahrheitsgemässe Berichterstattung und Teilnahmebedingung des Managements** Durch die spitalinterne asymmetrische Informationsverteilung zwischen Arzt und Management ändert sich für den Auftraggeber prinzipiell nichts. Auch hier muss er bei der Bestimmung von  $\{t^d(\theta), c^d(\theta)\}$  dafür sorgen, dass das Management die Fallschwere wahrheitsgemäss berichtet und dass das Management auch für jede Fallschwere die Behandlung genehmigt.

Allerdings muss der Auftraggeber beachten, dass das Management nun den „virtuellen“ Gewinn

$$\max_{\hat{\theta}^m} \tilde{G}(\theta, \hat{\theta}^m) = t^m(\hat{\theta}^m) + \theta V^*(c^m(\hat{\theta}^m)) - h(\theta)V^*(c^m(\hat{\theta}^m)) \quad (199)$$

anstatt des Gewinnes in Gleichung (168) als Grundlage für seine Entscheidungen betrachtet. Wenn also wahrheitsgemässe Berichterstattung die optimale Strategie des Managements ist, dann muss mit dem Envelope-Theorem und Gleichung (199)

$$\dot{\tilde{G}}(\theta) = (1 - h'(\theta))V^*(c^m(\theta)) > 0 \quad (200)$$

erfüllt sein.

Im Anhang A.5 wird gezeigt, dass die Anreizbedingung zusammen mit

$$\dot{c}^m(\theta) \geq 0 \quad (201)$$

hinreichend für wahrheitsgemässe Berichterstattung des Managements für alle Fallschweren ist.

Für den Pfad der Vergütung  $\dot{t}^m(\theta)$  entlang der Fallschwere ergibt sich mit Ableiten von (198) zusammen mit der Bedingung (200)

$$\dot{t}^m(\theta) = -\left(1 - \frac{h(\theta)}{\theta}\right)\lambda^*(\theta, \cdot)\dot{c}^m(\theta). \quad (202)$$

Damit das Management die Behandlung von Patienten jeder Fallschwere zulässt, muss der tatsächliche Nutzenüberschuss für jede Fallschwere mindestens den Reservationswert  $g$  erreichen. Der tatsächliche Nutzenüberschuss von Management und Arzt bei einem Patienten der Fallschwere  $\theta$  ist gegeben mit

$$N\ddot{U}(\theta) = t^m(\theta) + \theta V^*(c^m(\theta)) - u \quad (203)$$

Differenzieren nach  $\theta$  ergibt mit Einsetzen von (202) den „Pfad“ des Nutzenüberschusses, der aus der Anreizbedingung (200) folgt:

$$\dot{N}\ddot{U}(\theta) = \frac{h(\theta)}{\theta}\lambda^*(\theta, \cdot)\dot{c}^m(\theta) > 0. \quad (204)$$

Der Nutzenüberschuss muss also mit der Fallschwere ansteigen. Für die Teilnahmebedingung des Managements

$$N\ddot{U}(\theta) \geq g \quad \forall \theta \in \Theta \quad (205)$$

reicht es also aus, wenn der Nutzenüberschuss für die niedrigste Fallschwere mindestens bei  $g$  liegt:

$$N\ddot{U}(\underline{\theta}) \geq g. \quad (206)$$

Die Bedingung (206) impliziert, dass der Nutzenüberschuss für alle Fallschweren mindestens bei  $g$  liegt und das Management die Behandlung von Patienten aller Fallschweren



zulässt.

Da der Arzt für  $\underline{\theta}$  wegen (191) gerade seinen Reservationsnutzen erhält, lässt sich die Teilnahmebedingung des Managements auch als Mindestgewinnbedingung formulieren:

$$G(\underline{\theta}) \geq g. \quad (207)$$

### 8.2.2.5 Maximierungsproblem des Auftraggebers und optimale Allokation

Die Teilnahmebedingung (207) sowie die Anreizbedingung (200) implizieren einen Erwartungsgewinn für das Managements in Höhe von

$$E[G(\theta)] = \int_{\underline{\theta}}^{\bar{\theta}} \int_{\underline{\theta}}^{\theta} (1 - h'(\tilde{\theta})) V^*(c^m(\tilde{\theta})) d\tilde{\theta} f(\theta) d\theta + g. \quad (208)$$

Partielle Integration von (208) ergibt (vgl. (178))

$$E[G(\theta)] = \int_{\underline{\theta}}^{\bar{\theta}} h(\theta)(1 - h'(\theta)) V^*(c^m(\theta)) f(\theta) d\theta + g. \quad (209)$$

Mit (196) und (209) ergibt sich für den erwarteten Nutzenüberschuss  $EN\ddot{U} = E(U - u + G)^{68}$ :

$$E[N\ddot{U}(\theta)] = \int_{\underline{\theta}}^{\bar{\theta}} h(\theta)(2 - h'(\theta)) V^*(.) f(\theta) d\theta + g \quad (210)$$

Einsetzen von (210) in den Erwartungsnutzen des Auftraggebers

$$E[W(\theta)] = \int_{\underline{\theta}}^{\bar{\theta}} [\theta B^*(c^m(\theta)) + \theta V^*(c^m(\theta)) - c^m(\theta) - N\ddot{U}(\theta)] f(\theta) d\theta - u, \quad (211)$$

führt zu dessen Maximierungsproblem

$$\max_{c^m(\theta)} E[W(\theta)] = \int_{\underline{\theta}}^{\bar{\theta}} \theta (B^*(.) + V^*(.)) - c^m(\theta) - h(\theta)(2 - h'(\theta)) V^*(.) f(\theta) d\theta - g - u. \quad (212)$$

Wir gehen abermals davon aus, dass der Integrand von (212) für alle Fallschweren konkav ist, so dass die punktweise Maximierung des „virtuellen“ Nutzens des Sachwalters der Maximierung von (212) entspricht:

---

<sup>68</sup>Zur Erinnerung: Der interne Reservationsgewinn des Managements  $a$  wurde auf null gesetzt.

$$\max_{c^m(\theta)} \tilde{W} = \theta(B^*(.) + V^*(.)) - c^m(\theta) - h(\theta)(2 - h'(\theta)V^*(.)). \quad (213)$$

Die „virtuellen“ Kosten für den Auftraggeber belaufen sich jetzt auf  $h(\theta)(2 - h'(\theta)V^*(.))$ . Ein Vergleich mit der entsprechenden Gleichung im Modell ohne spitalinterne Informationsasymmetrie (180) zeigt, dass sich die „virtuellen“ Kosten für den Auftraggeber durch die spitalinterne asymmetrische Information mehr als verdoppelt haben, da  $-h'(\theta) > 0$ . Der Grund ist, dass nun die Information über die Fallschwere vom Auftraggeber im Prinzip zweimal „gekauft“ werden muss und zweimal Informationsrenten generiert werden, die ja letzten Endes vom Auftraggeber finanziert werden. Diese doppelte Rentengenerierung ist in der Literatur unter dem Stichwort „double marginalization of rents“ diskutiert [vgl. Melumad (1995)].

Wir gehen davon aus, (213) sei stetig und konkav in  $\theta$  und erhalten als Bedingung erster Ordnung für das optimale Fallbudget:

$$\left( \theta B_{q_1}^* \frac{\partial q_1^*(.)}{\partial c(\theta)} + \theta B_{q_2}^* \frac{\partial q_2^*(.)}{\partial c(\theta)} \right) + \left( 1 - \frac{h(\theta)(2 - h'(\theta))}{\theta} \right) \lambda^*(\theta, .) = 1. \quad (214)$$

Ein Vergleich der Bedingung (214) mit der entsprechenden Bedingung (181) des Modells ohne spitalinterne Informationsasymmetrie zeigt, dass sich die interne Informationsasymmetrie zwischen Arzt und Management in einer weiteren Verzerrung der optimalen Fallbudgetierung  $c^{m*}(\theta)$  nach unten bemerkbar macht. Zudem erhalten Management und Arzt zusammen einen höheren gemeinsamen Überschuss, da nun auch der Arzt in der Lage ist, eine Informationsrente zu generieren. Beläuft sich der gemeinsame Nutzenüberschuss von Management und Arzt bei vollständiger interner Information über die Fallschwere auf  $N\ddot{U}(\theta) = \int_{\underline{\theta}}^{\theta} V^*(c^m(\tilde{\theta}))d\tilde{\theta}$  (Gleichung (176)), so liegt er nun mit

$$N\ddot{U}(\theta) = \int_{\underline{\theta}}^{\theta} V^*(c^m(\tilde{\theta})) + \frac{h(\tilde{\theta})}{\tilde{\theta}} \lambda^*(\theta, .) \dot{c}^m(\tilde{\theta}) d\tilde{\theta} \quad (215)$$

für  $\theta > \underline{\theta}$  darüber. Lediglich bei einem Patienten der minimalen Fallschwere  $\underline{\theta}$  erhalten Arzt und Management den minimalen Überschuss  $g$ . Auf der anderen Seite resultiert wegen  $h(\bar{\theta}) = 0$  bei einem Patienten der höchsten Fallschwere dieselbe Budgetierung, die bei vollständiger Information über die Fallschwere optimal ist, jedoch muss der Auftraggeber Management und Arzt dann den grössten Nutzenüberschuss zugestehen. Insgesamt verstärkt die doppelte Rentengenerierung für den Auftraggeber den Trade-off zwischen effizienter Budgetierung und Informationsrente.

Allerdings ist auch im Falle von zusätzlicher spitalinterner Informationsasymmetrie hinsichtlich der Fallschwere nicht unbedingt gewährleistet, dass der optimale Pfad  $c^{m*}(\theta)$  die Bedingung zweiter Ordnung  $\dot{c}^m(\theta) \geq 0$  erfüllt, so dass auch hier eventuell für einen Intervall oder mehrere Intervalle der Fallschwere die Budgetierung konstant bleibt und eine „Bunching“-Problematik resultiert. Wir schliessen diesen Fall jedoch aus und fassen die Ergebnisse zusammen:

**Ergebnis 8.2** *Bei asymmetrischer Information bezüglich der Fallschwere zwischen Arzt und Management entstehen dem Management zusätzliche Kosten durch die Notwendigkeit, dem Arzt Anreize zur wahrheitsgemässen Berichterstattung der Fallschwere zu geben. Die Kosten entstehen zum einen aus der Informationsrente, die das Management dem Arzt zugestehen muss, zum anderen aus der Verzerrung der Berichtsstrategie, die eine effiziente interne Budgetierung verhindert. Diese Kosten werden nicht zwischen Arzt und Management in dem Sinne internalisiert, dass der gemeinsame Nutzenüberschuss von Arzt und Management davon unberührt bleibt. Vielmehr macht das Management die zusätzlichen Kosten gegenüber dem Auftraggeber geltend, wenn es den Bericht des Arztes hinsichtlich der Fallschwere wahrheitsgemäss an den Auftraggeber weiterleiten soll. Der gemeinsame Nutzenüberschuss von Arzt und Management bei spitalinterner asymmetrischer Information über die Fallschwere ist höher als der gemeinsame Nutzenüberschuss bei vollständiger interner Information. Für den Auftraggeber verstärkt sich damit der Trade-off zwischen Nutzenüberschuss und effizienter Budgetierung und die Fallbudgetierung erfährt eine weitere Verzerrung nach unten.*

Es ist offensichtlich, dass der Auftraggeber mit dem direkten Vergütungssystem, das bei vollständiger spitalinterner Information optimal ist, hier die optimale Allokation nicht erreichen kann, schon allein deswegen, da dieses System nicht die Anforderung der Anreizkompatibilität (204) erfüllt. In Bezug auf die Relevanz der spitalinternen Informationsverteilung bezüglich der Fallschwere folgern wir also:

**Ergebnis 8.3** *Um die optimale Allokation erreichen zu können, muss der Auftraggeber die spitalinternen Informationsverhältnisse bezüglich der Fallschwere kennen.*

### 8.3 Relevanz der Teilnahmebedingungen

Der zeitliche Ablauf des Modells lässt sich in Bezug auf den Informationsstand des Managements über die Fallschwere eines Patienten in zwei Abschnitte unterteilen, nämlich in den Zeitraum bevor der Arzt die Fallschwere im Rahmen der Aufnahmeuntersuchung des Patienten ermittelt und dem Management berichtet sowie in den Zeitraum danach. Der Auftraggeber gibt die Parameter  $t^m(\theta)$  und  $c^m(\theta)$  bekannt, bevor das Management die Fallschwere erfährt. Das Management muss also in Unkenntnis der tatsächlichen Fallschwere der einzelnen Patienten darüber entscheiden, ob es das Vergütungssystem des

Auftraggebers akzeptiert oder nicht. Dagegen fällt das Management die Entscheidung, ob es die Behandlung eines Patienten genehmigt oder nicht, erst nachdem es die Fallschwere des Patienten vom Arzt erfahren hat. Hinsichtlich des Ziels, mit seinem Vergütungs- und Budgetierungssystem die Behandlung eines Patienten zu erreichen, kann der Auftraggeber also an zwei Zeitpunkten scheitern: Zum einen, wenn das Management das Vergütungs- und Budgetierungssystem des Auftraggebers direkt nach der Bekanntgabe abweist, zum anderen, wenn das Management das System zum Zeitpunkt des Angebotes zwar akzeptiert, nach Ermittlung der Fallschwere aber die Behandlung Patienten nicht genehmigt.

Das heisst, der Auftraggeber muss bei der Bestimmung der freien Mittel  $t^m(\theta)$  und der Budgetierung  $c^m(\theta)$  aus zeitlicher Sicht die Erfüllung von zwei Teilnahmebedingungen einkalkulieren: Er muss zum einen dafür sorgen, dass das Management das System zum Zeitpunkt des Angebotes akzeptiert (in der zweiten Spielstufe des zeitlichen Ablauf des Modells, s. Abschnitt 8.1.4.1) und zum anderen muss er sicherstellen, dass das Management die Behandlung des Patienten nach Ermittlung der Fallschwere auch tatsächlich zulässt (in der dritten Spielstufe des zeitlichen Ablauf des Modells). In der Theorie wird die Teilnahmebedingung zum Zeitpunkt der Bekanntgabe des Vergütungssystems, an dem die Fallschwere noch nicht bekannt ist, als *ex-ante* Teilnahmebedingung bezeichnet. Die Teilnahmebedingung zum Zeitpunkt, an dem die Fallschwere bereits in Erfahrung gebracht wurde, wird dagegen als *ex-post* Teilnahmebedingung bezeichnet [Mas-Colell/Whinston/Green 1995, Seite 893].

Damit die *ex-ante* Teilnahmebedingung des Managements erfüllt ist, muss, da die Fallschwere noch nicht bekannt ist, der *erwartete* Nutzenüberschuss aus der Behandlung eines Patienten mindestens so hoch sein, wie der Wert, den die beste Alternative zum Angebot des Auftraggebers für das Management bedeutet. Damit hingegen die *ex-post* Teilnahmebedingung des Managements für alle Fallschweren erfüllt ist, muss, da das Management die Fallschwere nun kennt, der *ex-post* Nutzenüberschuss auch bei der aus Sicht des Managements ungünstigsten Fallschwere mindestens so hoch sein wie der Wert der besten Alternative zu diesem Zeitpunkt.

Wenn der Auftraggeber sein Vergütungs- und Budgetierungssystem *ex-post* nicht anpassen kann, was wir annehmen, dann muss er die Parameter  $t^m(\theta)$  und  $c^m(\theta)$  so bestimmen, dass die *ex-ante* Teilnahmebedingung *und* die *ex-post* Teilnahmebedingung des Managements nicht verletzt sind. Wie wir bereits gesehen haben, geht aber der Nutzenüberschuss von Management und Arzt  $N\check{U}(\theta)$  immer negativ in den Nutzen des Auftraggebers ein. Es ist also für den Auftraggeber immer optimal, den Nutzenüberschuss möglichst niedrig zu halten und ihn so weit nach unten zu „drücken“, bis entweder die *ex-ante* Teilnahmebedingung oder die *ex-post* Teilnahmebedingung für eine Fallschwere bindet. Die Frage ist,

welche der Teilnahmebedingungen es ist, die unter dem optimalen System zuerst bindet und welche Teilnahmebedingung damit automatisch erfüllt ist.

Bisher sind wir davon ausgegangen, dass eine (für die niedrigste Fallschwere) bindende ex-post Teilnahmebedingung des Managements

$$N\ddot{U}(\underline{\theta}) = g \quad (216)$$

hinreichend für die Erfüllung der ex-ante Teilnahmebedingung

$$\int_{\underline{\theta}}^{\bar{\theta}} N\ddot{U}(\theta) f(\theta) d\theta \geq \tilde{g} \quad (217)$$

des Managements ist, wobei  $\tilde{g}$  der Wert der besten Alternative pro Patient zum Zeitpunkt des Angebotes ist.

Jedoch ist auch der umgekehrte Fall denkbar, nämlich dass eine erfüllte ex-ante Teilnahmebedingung immer hinreichend für die Erfüllung der ex-post Teilnahme für alle Fallschweren ist. Hier ist es wichtig zu verstehen, dass sich die beste Alternative für das Management zum ex-ante Zeitpunkt von der besten Alternative zum ex-post Zeitpunkt unterscheidet und damit auch ihr Wert pro Fall. Entscheidend dafür, ob nun eine bindende ex-post Teilnahmebedingung hinreichend für eine Erfüllung der ex-ante Teilnahmebedingung ist oder umgekehrt, ist der Wert der Alternative zum ex-ante Zeitpunkt  $\tilde{g}$  im Verhältnis zum Wert zum ex-post Zeitpunkt  $g$ . Intuitiv ist einleuchtend, dass eine bindende ex-post Teilnahmebedingung automatisch die Teilnahme zum ex-ante Zeitpunkt bedeutet, wenn der Wert der Alternative zum ex-post Zeitpunkt  $g$  im Verhältnis zum Wert der Alternative zum ex-ante Zeitpunkt  $\tilde{g}$  besonders attraktiv ist.

Doch woraus besteht für das Management die Alternative zur Unterschrift unter das Angebot des Auftraggebers zum ex-ante Zeitpunkt und was bestimmt ihren Wert  $\tilde{g}$ ? Grundsätzlich bedeutet die Ablehnung des Angebotes, dass bezüglich der Behandlung der Patienten ein vertragsloser Zustand herrscht. Die Konsequenzen hieraus sind abhängig von ausserhalb des Modells definierten Rahmenbedingungen. Eine Alternative für das Management könnte darin bestehen, dem Arzt die Behandlung der entsprechenden Patienten zu verbieten, was ggf. zu erheblichen Umsatzausfällen führt, vor allem wenn das Management keinen anderen Vertrag mit einem anderen Auftraggeber abschliessen kann. Das Universitätsspital Zürich zum Beispiel generiert mit einem einzigen Vertrag, dem OKP-Vertrag, der die Vergütung der über die obligatorische Krankenversicherung gedeckten stationären Leistungen aller Zürcher Patienten regelt, über 50% seines gesamten stationären Umsatzes.

Bei einem solchem Vertrag ist es dem Management unmöglich, im Falle einer Ablehnung eine Alternative zu finden, die den essentiellen Umsatzausfall kompensieren könnte. Die Alternative zur Annahme ist also für das Management extrem unattraktiv. Eine andere Möglichkeit ist, dass übergeordnete Mechanismen greifen, wenn es zwischen Auftraggeber und Management zu keiner Einigung kommt, zum Beispiel eine behördliche Festsetzung. Hier ergibt sich der Wert der Alternative aus den Erwartungen bezüglich der behördlich festgesetzten Vergütung und Budgetierung.

Der Wert der Alternative zur ex-post Teilnahme des Managements ist ebenfalls vor allem durch die modellexogen definierten Rahmenbedingungen bestimmt, vor allem durch die rechtlichen. In diesem Zusammenhang wird häufig ein unter dem Stichwort „*Dumping*“ bekanntes Verhalten des Spitals in Theorie und Praxis diskutiert. Unter Dumping wird verstanden, wenn das Spital Patienten mit besonders unvorteilhaften Eigenschaften die Behandlung verweigert, zum Beispiel wenn die Behandlung besonders teuer und durch die Vergütung nicht gedeckt ist [vgl. Ellis (1998)]. Wenn zum Zeitpunkt des Vertragsabschlusses vereinbart wurde, dass jeder Patient unabhängig von seiner Fallschwere behandelt wird, dann stellt eine Abweisung eines Patienten einen Vertragsbruch dar, was unter Umständen zu nachteiligen Konsequenzen für das Management führt. Oder es besteht eine gesetzliche Aufnahmepflicht für das Spital, und eine Verweigerung der Aufnahme könnte ebenfalls empfindliche Konsequenzen für die Verantwortlichen nach sich ziehen, falls die Abweisung des Patienten nachzuweisen ist. So ist zum Beispiel im Gesundheitsgesetz des Kantons Zürich festgeschrieben, dass Spitäler mit kantonalem Leistungsauftrag Patienten, die einer stationären Behandlung bedürfen, aufnehmen müssen (KVG Art. 41a und Gesundheitsgesetz des Kantons Zürich, Aufnahmepflicht, §38,2). Wenn Dumping beobachtbar ist, dann stellt die Abweisung von Patienten also keine empfehlenswerte Alternative für das Management dar.<sup>69</sup>

Es wird deutlich, dass im Vertragsverhältnis zwischen Auftraggeber und Management durchaus Konstellationen denkbar sind, bei denen der Wert der Alternative zum ex-post Zeitpunkt im Verhältnis zur Alternative zum ex-ante Zeitpunkt so niedrig ist, dass eine erfüllte ex-ante Teilnahmebedingung hinreichend für die Erfüllung der ex-post Teilnahmebedingung für alle Fallschweren ist.

---

<sup>69</sup>Allerdings muss die Abweisung der 'falschen' Patienten nicht auf offensichtliche Art erfolgen. Das Spital könnte beispielsweise bestimmte Leistungen nicht oder nur in unzureichendem Ausmass anbieten, so dass gewisse Patienten erst gar nicht zur Aufnahme im Spital erscheinen. Diese Form der Patientenselektion ist natürlich schwerer zu beobachten und aus rechtlicher Sicht nicht unbedingt zu ahnden, so dass der Wert der Alternative eher durch die Möglichkeit der Umsatzsubstitution bestimmt wird und daher nicht allzu niedrig sein dürfte.

Entsprechendes gilt auch für das Verhältnis zwischen Arzt und Management. Für die Teilnahme des Arztes muss sowohl die entsprechende ex-ante sowie ex-post Bedingungen hinsichtlich seines Nutzens erfüllt sein. Wie beim Management unterscheiden sich auch für den Arzt die Werte der Alternativen zu den beiden Zeitpunkten, und auch hier gibt es eine Vielzahl von Faktoren, die den jeweiligen Wert der Alternative bestimmen. Lehnt der Arzt das vom Management festgelegte Paket aus freien Mitteln  $t^d(\theta)$  und Budgetierung  $c^d(\theta)$  zum ex-ante Zeitpunkt an sich ab, ist eine Option für den Arzt, das Spital zu verlassen. Der Wert dieser Option wird bei einem Arzt mit einem exzellenten medizinischen Ruf, dem anderswo vergleichbare Betätigungsfelder offen stehen, relativ hoch sein. Zum ex-post Zeitpunkt der Behandlung dagegen wird der Wert der Alternative wie für das Management von der Beobachtbarkeit von Dumping und den Regelungen bezüglich der Behandlungspflicht abhängen. Bei einem Arzt, der eine ausgewiesene Kapazität auf seinem Gebiet ist, wird also eher die ex-ante Teilnahmebedingung zu beachten sein, insbesondere wenn Dumping nicht erlaubt und leicht zu beobachten ist.

Wie bei Management sind wir bisher auch beim Arzt davon ausgegangen, dass eine bindende ex-post Teilnahmebedingung bei einem Patienten der niedrigsten Fallschwere

$$U(\underline{\theta}) = u \tag{218}$$

hinreichend für die Erfüllung der ex-ante Bedingung

$$\int_{\underline{\theta}}^{\bar{\theta}} U(\theta) f(\theta) d\theta \geq \tilde{u} \tag{219}$$

ist, wobei  $\tilde{u}$  der Wert der ex-ante Alternative pro Patient ist.

### 8.3.1 Lösung des Modells bei asymmetrischer Information auf beiden Ebenen und bei bindender ex-ante Teilnahmebedingung

Die Frage, ob nun die ex-ante oder die ex-post Teilnahmebedingung bindet, hat vor allem im Zusammenhang mit der der Problematik der versteckten Information zuzuordnenden Informationsasymmetrie hinsichtlich der Fallschwere Relevanz, da wegen der jeweiligen Anreizkompatibilitätsbedingungen von Arzt bzw. Management sich der ex-post Nutzen vom ex-ante Erwartungsnutzen bzw. der ex-post Gewinn vom Erwartungsgewinn in der Regel unterscheidet. Bei vollständiger Information hinsichtlich der Fallschwere ist der ex-post Nutzen bzw. der ex-post Gewinn identisch mit dem jeweiligen Erwartungswert.

Im Folgenden wollen wir analysieren, welchen Einfluss es auf die Ergebnisse des Modells bei asymmetrischer Information hinsichtlich der Fallschwere auf beiden Ebenen hat, wenn anstatt den ex-post Teilnahmebedingungen von Arzt und Management die jeweiligen ex-ante Teilnahmebedingungen relevant sind, zum Beispiel weil Dumping beobachtbar und strafbar ist, so dass der Wert der Option „Dumping“ zum ex-post Zeitpunkt so unattraktiv ist, dass sie für Arzt sowie Management nicht in Frage kommt.

Bei bindender ex-ante Teilnahmebedingung des Arztes (219) lautet das zu (193) äquivalente Maximierungsproblem des Managements

$$\max_{\hat{\theta}^m(\theta), U(\theta)} E[G(\theta)] = \int_{\underline{\theta}}^{\bar{\theta}} [t^m(\hat{\theta}^m(\theta)) + \theta V^*(c^m(\hat{\theta}^m(\theta)))] f(\theta) d\theta - \int_{\underline{\theta}}^{\bar{\theta}} U(\theta) f(\theta) d\theta \quad (220)$$

s.t.

$$\dot{U}(\theta) = V^*(c^m(\hat{\theta}^m(\theta))), \quad \int_{\underline{\theta}}^{\bar{\theta}} U(\theta) f(\theta) d\theta \geq \tilde{u}.$$

Mit der Anreizbedingung  $\dot{U}(\theta) = V^*(c^m(\hat{\theta}^m(\theta)))$  ergibt sich für den ex-post Nutzen des Arztes

$$U(\theta) = \int_{\underline{\theta}}^{\theta} V^*(c^m(\hat{\theta}^m(\tilde{\theta}))) d\tilde{\theta} + K, \quad (221)$$

wobei  $K$  eine vom Management zu bestimmende Konstante ist. Mit (221) lässt sich das Problem (220) alternativ formulieren als

$$\begin{aligned} \max_{\hat{\theta}^m(\theta), K} E[G(\theta)] &= \int_{\underline{\theta}}^{\bar{\theta}} [t^m(\hat{\theta}^m(\theta)) + \theta V^*(c^m(\hat{\theta}^m(\theta)))] f(\theta) d\theta - \\ &- \int_{\underline{\theta}}^{\bar{\theta}} \left[ \int_{\underline{\theta}}^{\theta} V^*(c^m(\hat{\theta}^m(\tilde{\theta}))) d\tilde{\theta} + K \right] f(\theta) d\theta \\ \text{s.t.} \\ &\int_{\underline{\theta}}^{\bar{\theta}} \left[ \int_{\underline{\theta}}^{\theta} V^*(c^m(\hat{\theta}^m(\tilde{\theta}))) d\tilde{\theta} + K \right] f(\theta) d\theta \geq \tilde{u}. \end{aligned} \quad (222)$$

Das Management bestimmt die Konstante  $K$  optimalerweise so niedrig, dass die Teilnahmebedingung des Arztes bindet, d.h.  $\int_{\underline{\theta}}^{\bar{\theta}} \left[ \int_{\underline{\theta}}^{\theta} V^*(c^m(\hat{\theta}^m(\tilde{\theta}))) d\tilde{\theta} + K \right] f(\theta) d\theta = \tilde{u}$ . Der Arzt erhält also keine ex-ante Informationsrente. Damit entspricht die Maximierung bezüglich des Berichtes  $\hat{\theta}^m(\theta)$  dem punktwisen Maximierungsproblem von



$$\max_{\hat{\theta}^m(\theta)} G(\theta) = t^m(\hat{\theta}^m(\theta)) + \theta V^*(c^m(\hat{\theta}^m(\theta))) - \tilde{u}. \quad (223)$$

Das Maximierungsproblem des Managements bei relevanter ex-ante Bedingung ist also (bis auf die für die Maximierung unbedeutende Konstante  $\tilde{u}$ ) identisch mit dem Maximierungsproblem (168) bei vollständiger spitalinterner Information. Insbesondere gilt die identische Bedingung für eine wahrheitsgemässe Berichterstattung des Managements:

$$\dot{G}(\theta) \equiv \frac{dG(\theta)}{d\theta} = V^*(c^m(\theta)) > 0. \quad (224)$$

Wir können also festhalten:

**Ergebnis 8.4** *Wenn die ex-ante Teilnahmebedingung des Arztes bindet, dann gilt bei spitalinterner asymmetrischer Information hinsichtlich der Fallschwere dieselbe Anreizbedingung des Managements gegenüber dem Auftraggeber, die auch bei vollständiger spitalinterner Information relevant ist. Der Arzt erhält eine erwartete Informationsrente von null.*

Damit das Management am Vertrag des Auftraggebers teilnimmt, muss der erwartete Überschuss mindestens bei  $\tilde{g}$  liegen:

$$\int_{\underline{\theta}}^{\bar{\theta}} N\ddot{U}(\theta) f\theta d\theta \geq \tilde{g}. \quad (225)$$

Das Maximierungsproblem des Auftraggebers lautet:

$$\max_{c^m(\theta), N\ddot{U}(\theta)} E[W(\theta)] = \int_{\underline{\theta}}^{\bar{\theta}} [\theta B^*(c^m(\theta)) + \theta V^*(c^m(\theta)) - c^m(\theta) - N\ddot{U}(\theta)] f\theta d\theta - \tilde{u}, \quad (226)$$

unter der Anreizbedingung (224) und der Teilnahmebedingung (225). Da der erwartete Nutzen des Arztes gerade bei  $\tilde{u}$  liegt, entspricht der erwartete Nutzenüberschuss dem erwarteten Gewinn des Managements,  $EN\ddot{U} = EG$ . Daher gilt mit der Anreizbedingung (224)

$$E[N\ddot{U}(\theta)] = \int_{\underline{\theta}}^{\bar{\theta}} \left[ \int_{\underline{\theta}}^{\theta} V^*(c^m(\hat{\theta}^m(\tilde{\theta}))) d\tilde{\theta} + A \right] f(\theta) d\theta \quad (227)$$

wobei  $A$  eine vom Auftraggeber zu bestimmende Konstante ist. Mit (227) lässt sich das Maximierungsproblem (226) des Auftraggebers äquivalent formulieren als

$$\begin{aligned} \max_{c^m(\theta), A} E[W(\theta)] &= \int_{\underline{\theta}}^{\bar{\theta}} [\theta B^*(c^m(\theta)) + \theta V^*(c^m(\theta)) - c^m(\theta)] f(\theta) d\theta - \\ &\quad - \int_{\underline{\theta}}^{\bar{\theta}} \left[ \int_{\underline{\theta}}^{\theta} V^*(c^m(\hat{\theta}^m(\tilde{\theta}))) d\tilde{\theta} + A \right] f(\theta) d\theta - \tilde{u}, \end{aligned} \quad (228)$$

unter der Teilnahmebedingung

$$\int_{\underline{\theta}}^{\bar{\theta}} \left[ \int_{\underline{\theta}}^{\theta} V^*(c^m(\hat{\theta}^m(\tilde{\theta}))) d\tilde{\theta} + A \right] f(\theta) d\theta \geq \tilde{g}. \quad (229)$$

Die Konstante  $A$  geht negativ in den Nutzen des Auftraggebers ein. Daher ist sie im Optimum so gewählt, dass die Teilnahmebedingung (229) gerade bindet. Das Management erhält keine ex-ante Informationsrente. Damit ist die Lösung des Problems (228) bezüglich der Budgetvorgabe  $c^m(\theta)$  identisch mit der Lösung des Problems

$$\max_{c^m(\theta)} \tilde{W}(\theta) = \theta B^*(c^m(\theta)) + \theta V^*(c^m(\theta)) - c^m(\theta) - \tilde{u} - \tilde{g} \quad (230)$$

für jede Fallschwere  $\theta$ . Das Problem (230) entspricht nun demjenigen bei vollständiger Information des Auftraggebers über die Fallschwere. Und damit ist auch die Bedingung erster Ordnung für den „Pfad“ der optimalen Budgetvorgabe

$$\left( \theta B_{q_1}^* \frac{\partial q_1^*(\cdot)}{\partial c(\theta)} + \theta B_{q_2}^* \frac{\partial q_2^*(\cdot)}{\partial c(\theta)} \right) + \lambda^*(\theta, \cdot) = 1. \quad (231)$$

identisch mit jener unter vollständiger Information bezüglich der Fallschwere (163).

**Ergebnis 8.5** *Wenn für die Teilnahme des Managements und des Arztes die jeweilige ex-ante Teilnahmebedingung bindet, dann ist aus Sicht des Auftraggebers der optimale Pfad der Fallbudgetierung identisch mit dem optimalen Pfad bei vollständiger Information über die Fallschwere. Arzt und Management erhalten beide eine erwartete Informationsrente von null, aus einer ex-ante Perspektive resultiert also dieselbe erwartete Allokation wie unter vollständiger Information hinsichtlich der Fallschwere.*

Für die Interpretation des Ergebnisses betrachten wir noch einmal das Verhältnis zwischen Management und Arzt: Ob nun die ex-ante Teilnahmebedingung oder die ex-post Teilnahmebedingung für die niedrigste Fallschwere des Arztes zuerst bindet, in jedem Fall muss das Management die Anreizbedingung (188) des Arztes beachten. Aus der Anreizbedingung des Arztes folgt, dass der Nutzen des Arztes mit der Fallschwere steigen muss.

Ist die *ex-post* Teilnahmebedingung des Arztes relevant, muss der Arzt für die Behandlung eines Patienten der niedrigsten Fallschwere mindestens den Reservationsnutzen  $u$  erhalten, für alle anderen Fallschweren muss der Nutzen wegen der Anreizbedingung aber darüber liegen. Damit ist auch der erwartete Nutzenüberschuss pro Patient grösser als  $g$ , so dass der Arzt auch aus der ex-ante Perspektive eine Informationsrente erhält. Dies führt dazu, dass das Management seinerseits von seiner Berichtstrategie abweicht, die es bei vollständiger spitalinterner Information verfolgt, um die Informationsrente zu verringern.

Ist dagegen die *ex-ante* Teilnahmebedingung des Arztes relevant, ist sein Nutzen aus der Behandlung eines Patienten wegen der Anreizbedingung ebenfalls dann am niedrigsten, wenn der behandelte Patient von der niedrigsten Fallschwere ist. Für die Teilnahme des Arztes reicht es jedoch aus, wenn der *erwartete* Nutzen pro Patient mindestens den Wert  $\tilde{u}$  erreicht. Ob der Nutzen aus der Behandlung bei einzelnen Fallschweren unterhalb  $\tilde{u}$  liegt, spielt dabei keine Rolle. Mit der geeigneten Wahl des Pfades der freien Mittel  $t^d(\theta)$  kann das Management bei *gegebenem* Pfad der Berichtstrategie  $\hat{\theta}^m(\theta)$  erreichen, dass der *erwartete* Nutzen des Arztes gerade bei  $\tilde{u}$  liegt und der Arzt aus einer ex-ante Perspektive keine Informationsrente generieren kann. Insbesondere gilt dies auch für den unverzerrten Pfad der Berichtstrategie, der Pfad also, der optimal ist, wenn zwischen Management und Arzt keine asymmetrische Information bezüglich der Fallschwere herrscht.

Da wegen der Risikoneutralität von Management und Arzt gegenüber der freien Mittel jede Allokation der freien Mittel ohne Effizienzverlust ist, wählt das Management den verzerrungsfreien Pfad der Berichterstattung und erreicht die Erfüllung der Anreizbedingung und der Teilnahmebedingung des Arztes mit dem dafür notwendigen Pfad der freien Mittel  $t^d(\theta)$ . Daher resultiert bei relevanter ex-ante Teilnahmebedingung des Arztes dieselbe Anreizbedingung des Managements, die für den Auftraggeber zu beachten ist, wenn zwischen Arzt und Management keine asymmetrische Information bezüglich der Fallschwere vorliegt.

Für das Verhältnis zwischen Auftraggeber und Management gilt ein entsprechender Zusammenhang. Um den wahrheitsgemässen Bericht des Managements zu erhalten, muss der Auftraggeber bei relevanter ex-ante Teilnahmebedingung genauso wie bei relevanter ex-post Teilnahmebedingung die Anreizbedingung des Managements (169) beachten. Aus der Anreizbedingung folgt, dass der Nutzenüberschuss von Management und Arzt mit der Fallschwere steigen muss.

Ist die *ex-post* Teilnahmebedingung relevant, erhält das Management für alle Fallschweren ausser der niedrigsten eine Informationsrente grösser null, so dass auch im Erwartungswert eine positive Informationsrente resultiert. Ist dagegen die *ex-ante* Teilnahmebedin-

gung des Managements relevant, ist es dem Auftraggeber möglich, bei *gegebenem* Pfad der Fallbudgetierung  $c^m(\theta)$  mit der geeigneten Wahl des Pfades der Vergütung  $t^m(\theta)$  die Anreizbedingung des Management zu erfüllen und dem Management einen ex-ante Nutzenüberschuss zuzugestehen, der gerade beim Reservationswert  $\tilde{g}$  liegt. Da wegen der Risikoneutralität von Auftraggeber und Management die Höhe der Vergütung die Effizienz nicht beeinträchtigt, kann der Auftraggeber den Pfad der Budgetierung so bestimmen, dass der verzerrende Effekt der asymmetrischen Information eliminiert wird.

Zur Verdeutlichung des Sachverhaltes ist im oberen Teil von Abbildung (3) der Verlauf der Fallbudgetierung bei relevanter ex-ante Teilnahmebedingung (durchgezogene Linie) und bei relevanter ex-post Teilnahmebedingung (gestrichelte Linie) dargestellt, unter der Annahme, dass beim Arzt die ex-ante Teilnahmebedingung relevant ist. Die Budgetierung ist für die höchste Fallschwere identisch, für alle anderen Fallschweren ist sie im Falle der ex-post Teilnahmebedingung nach unten verzerrt. Der untere Teil der Abbildung enthält die Darstellung des jeweiligen Nutzenüberschusses des Managements. Zur Vereinfachung gilt in der Darstellung  $g = \tilde{g}$ . Der erwartete Nutzenüberschuss bei relevanter ex-ante Bedingung liegt bei  $\tilde{g}$ , gilt dagegen die ex-post Teilnahmebedingung, liegt der erwartete Nutzenüberschuss darüber,  $EN\ddot{U} > \tilde{g}$ .

## 8.4 Einfluss der Nash-Verhandlungslösung auf das Ergebnis

Zum Abschluss der Betrachtung wollen wir - auch im Sinne der Vollständigkeit - kurz darauf eingehen, welchen Einfluss es hat, wenn bei asymmetrischer Information hinsichtlich der Fallschwere die interne PA-Lösung durch die NV-Lösung ersetzt wird. Wir beschränken uns dabei allerdings nur auf die Überprüfung, ob die Äquivalenz der PA-Lösung und der NV-Lösung hier Bestand hat oder nicht.

Wir gehen von derselben zeitlichen Abfolge wie in Abschnitt 6.3 aus, d.h. Arzt und Management verhandeln die freien Mittel  $t^d(\theta)$  und die Budgetierung  $c^d(\theta)$  *nach* der Bekanntgabe des Vergütungs- und Budgetierungssystems  $\{t^m(\theta), c^m(\theta)\}$  durch den Auftraggeber, aber *vor* der Behandlung des Patienten und Beobachtung der Fallschwere durch den Arzt. Da der Auftraggeber jedoch das direkte System  $\{t^m(\theta), c^m(\theta)\}$  einsetzt, ist die Verhandlung einer Budgetierung  $c^d(\theta)$  gleichbedeutend mit der Verhandlung der Berichtstrategie  $\hat{\theta}^m(\theta)$  des Managements. Arzt und Management klären also in ihrer Verhandlung vorab, wie das Management den späteren Bericht des Arztes gegenüber dem Auftraggeber verwerten wird.

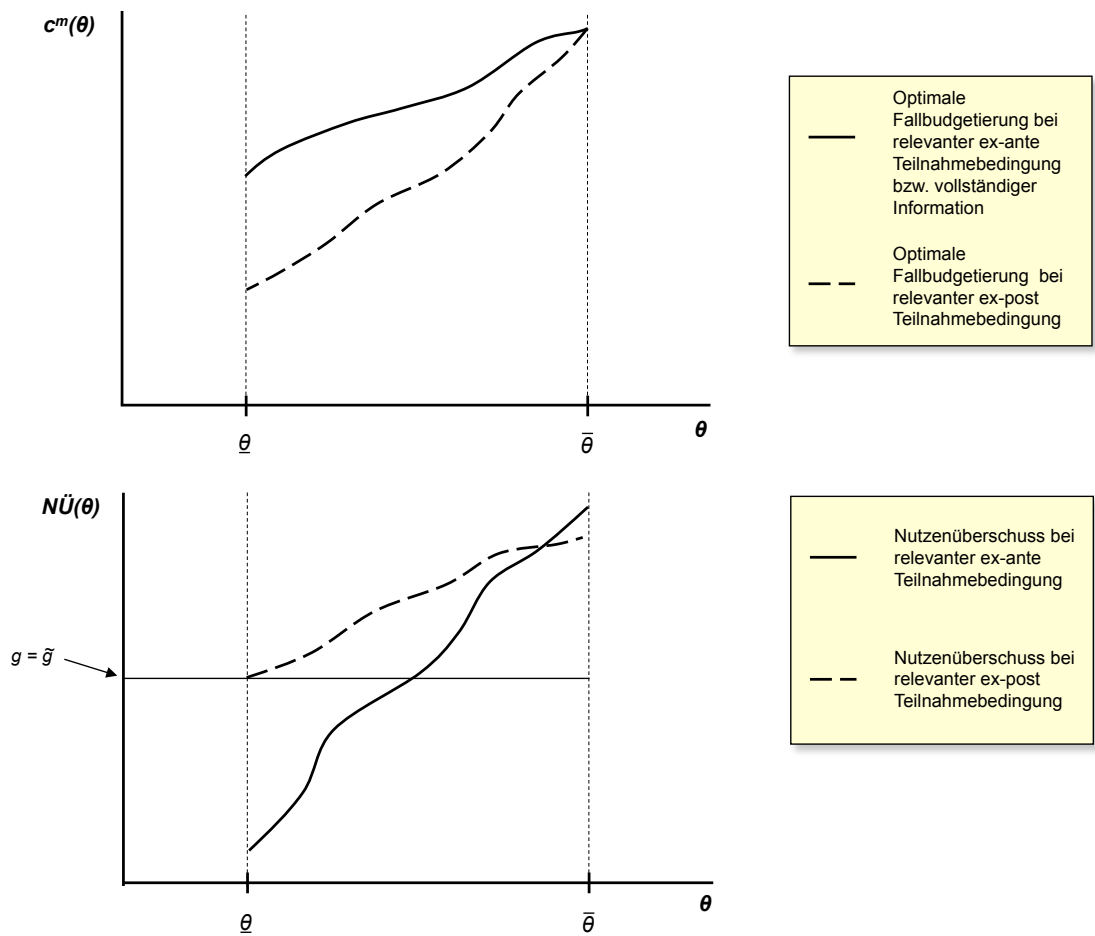


Abbildung 3: Verlauf der Fallbudgetierung und des Nutzenüberschusses bei relevanter ex-ante Teilnahmebedingung (durchgezogene Linie) und ex-post Teilnahmebedingung (gestrichelte Linie).

#### 8.4.1 Bedeutung der Verhandlungslösung bei vollständiger spitalinterner Information hinsichtlich der Fallschwere zwischen Arzt und Management

Zunächst betrachten wir die Situation, in der zwischen Arzt und Management keine asymmetrische Information hinsichtlich der Fallschwere vorliegt. Die Nash-Verhandlungslösung ist in diesem Falle die Lösung des Maximierungsproblems

$$\max_{t^d(\theta), \hat{\theta}^m(\theta)} \pi(\theta) = G(\theta)^w (U(\theta) - u)^{1-w}. \quad (232)$$

Wir wissen, dass die Nash-Verhandlungslösung auf der Pareto-effizienten Grenze liegt. Da wir Risikoneutralität von Management und Arzt annehmen sowie davon ausgehen, dass Arzt und Management beide die freien Mittel mit dem Gewicht 1 bewerten, ist die Pareto-effiziente Grenze eine Gerade mit der Steigung  $-1$ , d.h. der Nutzenüberschuss

$$N\ddot{U}(\theta, \hat{\theta}^m(\theta)) = t^m(\hat{\theta}^m(\theta)) + \theta V^*(c^m(\hat{\theta}^m(\theta))) - u \quad (233)$$

hat für alle Punkte auf der Pareto-effizienten Grenze den gleichen Wert. Daher maximiert die Lösung von (232) bezüglich der Berichtstrategie  $\hat{\theta}^m(\theta)$  auch das Problem

$$\max_{\hat{\theta}^m(\theta)} N\ddot{U}(\theta, \hat{\theta}^m(\theta)) = t^m(\hat{\theta}^m(\theta)) + \theta V^*(c^m(\hat{\theta}^m(\theta))) - u. \quad (234)$$

Das Problem (234) ist identisch mit dem des Managements (168) der PA-Lösung bei vollständiger spitalinternen Information. Als Bedingung für wahrheitsgemäße Berichterstattung erhalten wir daher mit dem Envelope-Theorem (vgl. (169))

$$\dot{N\ddot{U}}(\theta) = V^*(c^m(\theta)) \quad \forall \theta \in \Theta. \quad (235)$$

Wenn die Bedingung (235) (zusammen mit der Bedingung zweiter Ordnung (170)) erfüllt ist, dann einigen sich Arzt und Management auf eine Lösung, die wahrheitsgemäße Berichterstattung durch das Management impliziert.

Aus Sicht des Auftraggebers ist die Bedingung für wahrheitsgemäße Berichterstattung im Falle der NV-Lösung (235) identisch mit der entsprechenden Bedingung im Falle der PA-Lösung (169). Auch die Teilnahmebedingung des Managements ist mit derjenigen bei der PA-Lösung identisch:

$$N\ddot{U}(\underline{\theta}) \geq g. \quad (236)$$

Der Auftraggeber löst also exakt dasselbe Maximierungsproblem wie im Falle der PA-Lösung und wir folgern in Bezug auf die Äquivalenz der PA-Lösung und der NV-Lösung:

**Ergebnis 8.6** *Wenn lediglich zwischen Management und Auftraggeber, nicht aber zwischen Arzt und Management asymmetrische Information hinsichtlich der Fallschwere besteht, dann ist die NV-Lösung zur PA-Lösung äquivalent.*

#### 8.4.2 Einfluss der Verhandlungslösung bei asymmetrischer Information hinsichtlich der Fallschwere zwischen Arzt und Management

Eine andere Situation ergibt sich, wenn man davon ausgeht, dass der Arzt die Fallschwere nur dann wahrheitsgemäss dem Management berichtet, wenn dies für ihn optimal ist. Denn der Umstand, dass der Arzt die freien Mittel  $t^d(\theta)$  und die Berichtsstrategie  $\hat{\theta}^m(\theta)$  für alle Fallschweren im Rahmen der Verhandlungen mit dem Management in einem entscheidenden Mass *ex-ante* mitbestimmt, bedeutet nicht automatisch, dass der Arzt *ex-post* wahrheitsgemäss berichtet. Genauso gut könnte auch eine Lösung ausgehandelt werden, die den *ex post* Nutzen des Arztes bei einer falschen Berichterstattung maximiert.

Allerdings lässt sich mit dem Prinzip der direkten Offenbarung plausibel begründen, dass Arzt und Management eine Lösung verhandeln, die den Arzt später zu wahrheitsgemässer Berichterstattung veranlasst: Angenommen, Arzt und Management verhandeln freie Mittel und eine Berichtsstrategie  $\{t^{d'}(\theta), \hat{\theta}^{m'}(\theta)\}$  für alle Fallschweren so, dass der Arzt später die Fallschwere nicht wahrheitsgemäss berichtet und eine entsprechende Allokation resultiert. Nach dem Prinzip der direkten Offenbarung gibt es dann eine andere Verhandlungslösung  $\{t^{d''}(\theta), \hat{\theta}^{m''}(\theta)\}$  für alle Fallschweren, welche dieselbe Allokation implementiert, den Arzt jedoch zu wahrheitsgemässer Berichterstattung veranlasst. Gegenüber dieser Lösung ziehen Arzt und Management jedoch die Allokation (schwach) vor, die aus der optimalen Verhandlungslösung  $\{t^{d*}(\theta), \hat{\theta}^{m*}(\theta)\}$  für alle Fallschweren resultiert.

Wir gehen also davon aus, dass sich Management und Arzt bei den Verhandlungen des Problems der späteren wahrheitsgemässen Berichterstattung bewusst sind und wissen, dass die optimale Lösung  $\{t^{d*}(\theta), \hat{\theta}^{m*}(\theta)\} \forall \theta \in \Theta$  den Arzt zu wahrheitsgemässer Berichterstattung veranlassen muss, d.h. dass die Lösung die Bedingung des Arztes für wahrheitsgemässe Berichterstattung

$$\dot{U}(\theta) = V^*(c^m(\hat{\theta}^{m*}(\theta))) \forall \theta \in \Theta \quad (237)$$

erfüllen muss (vgl. (188)).

Die Bedingung (237) impliziert, dass der Arzt eine *ex-post* Informationsrente erhalten muss, die mit der Fallschwere steigt. Bei der Verhandlungslösung werden die optimalen Pfade der freien Mittel  $t^{d*}(\theta)$  und der optimalen Berichtsstrategie  $c^{d*}(\theta)$  verhandelt, und

der gemeinsame Nutzenüberschuss den spitalinternen Machtverhältnissen entsprechend zwischen Arzt und Management verteilt. Im Anteil des Arztes am gemeinsamen Nutzenüberschuss ist die Informationsrente des Arztes jedoch enthalten und da es für den Arzt keine Rolle spielt, ob sein Anteil am Nutzenüberschuss aus der Verhandlungslösung oder aus der Informationsrente stammt, hat die Informationsrente des Arztes nicht den verzerrenden Effekt, den sie bei der PA-Lösung hat. Unter Umständen ist es bei der NV-Lösung sogar möglich, dass die Informationsrente des Arztes vollständig spitalintern internalisiert wird, und die Verhandlungslösung den gemeinsamen Nutzenüberschuss maximiert.

In Anhang A.6 wird gezeigt, dass dies tatsächlich der Fall ist, wenn der Arzt über eine gewisse Verhandlungsmacht gegenüber dem Management verfügt und der Auftraggeber das Spital mit demselben System vergütet, das bei vollständiger spitalinterner Information hinsichtlich der Fallschwere optimal ist. In diesem Fall ist die spitalinterne asymmetrische Information hinsichtlich der Fallschwere für den Auftraggeber ohne Bedeutung und wahrheitsgemässe Berichterstattung durch das Management bzw. Spital maximiert trotz der Bedingung (237) den gemeinsamen Nutzenüberschuss von Arzt und Management, d.h. die verhandelte Berichtstrategie  $\hat{\theta}^{m*}(\theta) = \theta$  ist die Lösung des Maximierungsproblems

$$\max_{\hat{\theta}^m(\theta)} N\ddot{U}(\theta, \hat{\theta}^m(\theta)) = t^m(\hat{\theta}^m(\theta)) + \theta V^*(c^m(\hat{\theta}^m(\theta))) - u \quad (238)$$

für jedes  $\theta \in \Theta$ .

Zur PA-Lösung hingegen ist die NV-Lösung bei spitalinterner Informationsasymmetrie hinsichtlich der Fallschwere nicht äquivalent, da bei der PA-Lösung die Informationsrente des Arztes in keinem Fall internalisiert wird. Wir halten also als Ergebnis fest:

**Ergebnis 8.7** *Bei spitalinterner asymmetrischer Information hinsichtlich der Fallschwere ist die Äquivalenz der NV-Lösung zur PA-Lösung nicht gegeben. Unter Umständen wird bei der NV-Lösung die Informationsrente des Arztes spitalintern vollständig internalisiert, so dass die spitalinterne Informationsasymmetrie für den Auftraggeber ohne Bedeutung ist.*

## 8.5 Fazit Kapitel 8

Die Analyse in diesem Kapitel zeigt, dass es im Modell mit doppeltem Unterstellungsverhältnis durchaus eine Rolle spielt, ob es spitalintern zwischen Arzt und Management zu asymmetrischer Information hinsichtlich der Fallschwere kommt oder nicht.



Wenn hinsichtlich der Fallschwere spitalintern *vollständige* Information vorliegt, gesteht das Management dem Arzt bei der Behandlung jedes Patienten gerade seinen Reservationsnutzen zu, unabhängig von der Fallschwere. Auf der anderen Seite kann das Management aus dem Informationsvorsprung hinsichtlich der Fallschwere gegenüber dem Auftraggeber eine Informationsrente erlangen. Diese Informationsrente kann der Auftraggeber verringern, muss dafür aber eine Verzerrung der optimalen Budgetvorgabe nach unten hinnehmen.

Besteht dagegen auch spitalintern asymmetrische Information über die Fallschwere, kommt es zu einer doppelten Generierung von Informationsrenten: Wenn zunächst nur der Arzt über die Fallschwere informiert ist, kann er diesen Informationsvorsprung in einer Informationsrente umsetzen. Da diese Informationsrente zulasten des Managements geht, versucht das Management sie zu minimieren, muss dabei aber von der optimalen Strategie der Weiterleitung der vom Arzt berichteten Fallschwere an den Auftraggeber abweichen. Der zusätzliche Nutzen für den Arzt wird jedoch nicht zwischen Arzt und Management internalisiert, im Sinne einer Aufteilung eines konstanten gemeinsamen Nutzenüberschusses. Vielmehr überwälzt das Management die zusätzlichen Kosten an den Auftraggeber. Zusätzlich kann das Management eine „eigene“ Informationsrente generieren, da es nach dem Bericht des Arztes gegenüber dem Auftraggeber einen Informationsvorsprung erlangt. Im Vergleich zur Situation bei vollständiger spitalinterner Information resultiert eine weitere Verzerrung der optimalen Fallbudgetierung nach unten.

Der Auftraggeber benötigt also Informationen über die spitalinterne Situation der Informationsverteilung, um das optimale Vergütungssystem zu bestimmen. Dies gilt auch dann, wenn das Abweisen von Patienten sehr unattraktiv für das Management bzw. den Arzt ist und dafür die jeweilige ex-ante Teilnahmebedingung relevant ist. Zwar resultiert für jede Fallschwere dasselbe Budget wie bei vollständiger Information, trotzdem bleibt der Auftraggeber auf die wahrheitsgemässe Berichterstattung von Arzt und Management angewiesen und muss die Vergütung entsprechend der Anreizbedingung des Managements gestalten. Bei vollständiger spitalinterner Information unterscheidet sich diese jedoch von der jener bei spitalinterner asymmetrischer Information hinsichtlich der Fallschwere.

Bezüglich der Äquivalenz der NV-Lösung zur PA-Lösung ergibt sich, dass der interne Entscheidungsmechanismus für den Auftraggeber dann keine Rolle spielt, wenn auch das Management vollständig über die Fallschwere informiert ist. Bei spitalinterner asymmetrischer Information über die Fallschwere hingegen ist die Äquivalenz der beiden Mechanismen nicht gegeben, da die Informationsrente des Arztes spitalintern bei der PA-Lösung anders bewertet wird als unter der NV-Lösung. Wenn spitalintern asymmetrische Information hinsichtlich der Fallschwere vorliegt, dann benötigt er zusätzlich die Information,

ob es zwischen Management und Arzt zu einer PA-Lösung oder einer NV-Lösung kommt.

## 9 Schlussfolgerung

### 9.1 Relevanz des spitalinternen Entscheidungsprozesses für das optimale Vergütungssystem

Ziel dieser Arbeit war es, den Einfluss verschiedener spitalinterner Entscheidungsprozesse auf das Verhalten des Spitals gegenüber dem Vergütungssystem des Auftraggebers bzw. den darin enthaltenen Anreizen zu analysieren. Hierfür wurde ein erweitertes PA-Modell entwickelt, in dem ein Auftraggeber die Behandlung von Patienten an ein Spital delegiert. Im Modell galt die Annahme, dass spitalintern die für die Interaktion mit dem Auftraggeber relevanten Entscheidungen vom Spitalmanagement und vom Arzt getroffen werden, wobei grundsätzlich angenommen wurde, dass dem Arzt die Behandlung der Patienten obliegt, dem Spitalmanagement dagegen die Organisation des Spitals und die Kommunikation mit dem Auftraggeber. Die Allokation der Ressourcen zwischen Arzt und Management resultiert im Modell jedoch aus einem konkreten, spitalinternen Entscheidungsprozess, so dass das Verhalten des Spitals im Modell einerseits von den Präferenzen von Arzt und Management abhängt, andererseits aber von dem angenommenen spitalinternen Entscheidungsprozess, der den Weg der Entscheidungsfindung beschreibt.

Im Modell ist der spitalinterne Entscheidungsprozess durch zwei Aspekte gekennzeichnet. Erstens durch den Mechanismus der Entscheidungsfindung, der die formellen Regeln der Interaktion zwischen Arzt und Management definiert und die Kompetenzen zur Festlegung bestimmter Grössen vergibt, und zweitens durch Informationsasymmetrien zwischen Arzt und Management, welche die vertragliche Festlegung der entsprechenden Grössen spitalintern verhindert. Beim Mechanismus der Entscheidungsfindung wurde zum einen der Fall betrachtet, bei dem die Fallbudgetierung, die der Arzt bei der Behandlung eines Patienten einhalten muss, sowie die finanzielle Ausstattung der klinischen Abteilung des Arztes, die über die Deckung der Behandlungskosten hinausgeht (freie Mittel), vom Management als Prinzipal des Arztes autonom bestimmt werden (PA-Lösung). Zum anderen wurde als alternativer Mechanismus eine Situation angenommen, in welcher die Fallbudgetierung und die freien Mittel als Nash-Verhandlungslösung zwischen Management und Arzt bestimmt werden, in der sich die Verhandlungsmacht des Arztes gegenüber dem Management widerspiegelt.

Hinsichtlich der asymmetrischen Informationsverteilung zwischen Auftraggeber und Spital wurden in den Kapiteln 6, 7 und 8 verschiedene Szenarien betrachtet, innerhalb dieser Szenarien wurden verschiedene Entscheidungsprozesse miteinander verglichen. Insbesondere wurde der Frage nachgegangen, ob diese zu unterschiedlichem Verhalten des Spitals gegenüber dem Vergütungssystem des Auftraggebers führen, oder ob sie aus Sicht des

Auftraggebers äquivalent sind. Bezüglich dieser Fragestellung wurde in der Analyse deutlich, dass es hinreichend für die Äquivalenz zweier Entscheidungsprozesse ist, wenn zum einen aus beiden Prozessen eine Allokation resultiert, die zwischen Arzt und Management Pareto-effizient ist, und zum anderen die absolute Höhe des gemeinsamen Nutzenüberschusses von Arzt und Management unabhängig von der Verteilung des Überschusses unter den beiden ist. Unter diesen Umständen ist die effiziente spitalinterne Budgetierung bei beiden Mechanismen identisch und maximiert den gemeinsamen Nutzenüberschuss unabhängig von der Verteilung desselben. Die aus den beiden verschiedenen Prozessen resultierenden Allokationen unterscheiden sich also höchstens in der Verteilung des gemeinsamen Nutzenüberschusses. Da aber der Auftraggeber grundsätzlich indifferent gegenüber der spitalinternen Aufteilung des Nutzenüberschusses ist, sind aus seiner Sicht beide Prozesse äquivalent, zumindest solange die spitalinternen Teilnahmebedingungen von Arzt und Management nicht verletzt sind.

In Kapitel 6 wurde eine Situation betrachtet, in der die zur Behandlung eines Patienten eingesetzten Mengen der medizinischen Leistungen private Information des Arztes sind. Wenn Management und Auftraggeber risikoneutral gegenüber den freien Mitteln sind, dann spielt es in dieser Situation für den Auftraggeber keine Rolle, ob das Management die Mengen beobachten kann oder nicht, und ob zwischen Management und Arzt ein PA-Verhältnis vorliegt oder die Budgetierung und die freien Mittel eine Nash-Verhandlungslösung zwischen Arzt und Management darstellen. Die resultierende Allokation ist bei allen Entscheidungsprozessen Pareto-effizient. Die asymmetrische Information spielt keine Rolle, da das Management per Annahme indifferent hinsichtlich zur Behandlung eingesetzten Mengen ist, solange die Budgetvorgabe eingehalten wird. Von daher sind die vom Arzt gewählten Mengen aus Sicht des Managements effizient, da sie bei gegebener Budgetrestriktion den Nutzenüberschuss maximieren. Da die PA-Lösung und die Nash-Verhandlungslösung beide Pareto-effizient sind und der Nutzenüberschuss „verlustfrei“ über die freien Mittel zwischen Arzt und Management aufgeteilt werden kann, resultiert bei beiden Mechanismen die gleiche Budgetierung, die bei gegebenem Vergütungssystem des Auftraggebers den gemeinsamen Nutzenüberschuss maximiert. Bei der Nash-Verhandlungslösung beeinflusst die Verhandlungsmacht des Arztes nur die Höhe der freien Mittel und damit die Verteilung des Nutzenüberschusses, nicht aber die Budgetierung. Da für den Auftraggeber aber nur die resultierende Budgetierung relevant ist, sind für ihn unter den gegebenen Umständen die beiden Entscheidungsmechanismen äquivalent.

Ist allerdings bei gegebener Vergütung durch den Auftraggeber eine Verteilung des gemeinsamen Nutzenüberschusses zwischen Management und Arzt nicht möglich, ohne dass dadurch der gemeinsame Nutzenüberschuss in seiner Höhe beeinflusst wird, ist die Äqui-

valenz zwischen PA-Lösung und NV-Lösung nicht gegeben. Ist zum Beispiel der Arzt risikoavers gegenüber den freien Mitteln, ist die Umverteilung des Nutzenüberschusses über die freien Mittel wegen des abnehmenden Grenznutzen des Arztes nicht verlustfrei möglich. Der vom Arzt unter der NV-Lösung in Anspruch genommene Anteil am Nutzenüberschuss wird dann nur zu einem Teil über die freien Mittel finanziert. Der Rest wird mit einer vergleichsweise höheren Budgetierung ermöglicht. Da der Anteil des Arztes am gemeinsamen Nutzenüberschuss umso höher ausfällt, je grösser seine Verhandlungsmacht gegenüber dem Management ist, hängt die Budgetierung in diesem Fall von der Verteilung der Verhandlungsmacht ab.

In Kapitel 7 wurde die in Kapitel 6 betrachtete Situation um den Aspekt der Kosteneffizienz erweitert. Es wurde angenommen, dass das Management durch entsprechende Massnahmen den Preis für die Mengen  $q_1$  beeinflussen kann, bevor die spitalinterne Budgetierung festgelegt wird. Auch mit dieser Erweiterung ist die PA-Lösung Pareto-effizient, da das Management bei der Bestimmung des Budgets das ihm durch die Massnahmen zur Senkung des Preises  $p_1$  entstandene Arbeitsleid natürlich vollständig berücksichtigt, d.h. die Anstrengungen werden bei der Bestimmung des Preises und der Budgetierung vollständig zwischen Management und Arzt internalisiert. Die NV-Lösung ist dagegen nur dann Pareto-effizient, wenn das Management bei der Festlegung der Massnahmen davon ausgehen kann, dass es das ihm dadurch entstehende Arbeitsleid in der späteren Verhandlung mit dem Arzt geltend machen kann. Ist ihm dies nur zum Teil möglich, führt das Management die Massnahmen weniger intensiv durch als für eine Pareto-effiziente Lösung notwendig wäre. Dieser Effekt ist um so ausgeprägter, je grösser die Verhandlungsmacht des Arztes gegenüber dem Management ist. In dieser Situation ist also auch bei risikoneutralem Arzt die Verhandlungsmacht des Arztes von Bedeutung für das optimale Vergütungssystem und die NV-Lösung ist nicht äquivalent zur PA-Lösung.

In Kapitel 8 wurde mit der asymmetrischen Information über die Fallschwere eine Problematik der versteckten Information thematisiert. Hier versucht der Auftraggeber mit einer Anreizstruktur das Management dazu zu bewegen, die Fallschwere des behandelten Patienten wahrheitsgemäss zu offenbaren. Besteht lediglich zwischen Management und Auftraggeber asymmetrische Information hinsichtlich der Fallschwere, nicht jedoch zwischen Management und Arzt, dann sind die PA-Lösung und die NV-Lösung äquivalent. In beiden Fällen resultiert die gleiche Strategie der Berichterstattung der Fallschwere, welche zu einer Budgetierung führt, die den gemeinsamen Nutzenüberschuss maximiert. Auch hier unterscheiden sich die beiden Lösungen nur in der Höhe der freien Mittel für den Arzt. Herrscht dagegen auch spitalintern zwischen Arzt und Management asymmetrische Information hinsichtlich der Fallschwere, ist weder die PA-Lösung noch die NV-Lösung zwischen Management und Arzt Pareto-effizient. Besteht zwischen Management und Arzt

ein PA-Verhältnis, muss das Management einen Trade-off zwischen effizienter Budgetierung (bzw. effizienter Berichterstattung) und einer Informationsrente für den Arzt eingehen, welche der Arzt aufgrund seines Informationsvorsprunges generieren kann. Wird die Budgetierung (bzw. die Strategie der Berichterstattung) zwischen Arzt und Management verhandelt, wird bei dieser Entscheidungsfindung die Informationsrente des Arztes allerdings anders bewertet, als wenn das Management die Budgetierung autonom festlegt. Aus Sicht des Auftraggebers resultiert also bei spitalinterner NV-Lösung eine andere Bedingung für wahrheitsgemäße Berichterstattung als bei spitalinternem PA-Verhältnis zwischen Management und Arzt.

## **9.2 Umsetzung des optimalen Systems: Nutzengewinn und Kosten für den Auftraggeber**

Es lässt sich also festhalten, dass im Rahmen des Modells Informationen über die spitalinternen Gegebenheiten durchaus relevant für den Auftraggeber sind, wenn er das im Sinne einer second best-Lösung optimale Vergütungssystem bestimmen will. Aus der Analyse geht allerdings nicht hervor, ob sich die *Umsetzung* des optimalen Systems tatsächlich lohnt, da die Kosten der Bestimmung und die Implementierung eines optimalen Vergütungssystems für den Auftraggeber nicht berücksichtigt wurden. Tatsächlich dürften diese Kosten jedoch erheblich sein. So muss der Auftraggeber ggf. nicht nur darüber informiert sein, ob spitalintern die relevanten Entscheidungen im Rahmen von Verhandlungen getroffen werden, sondern er muss darüber hinaus auch die Verteilung der Verhandlungsmacht zwischen Management und Arzt abschätzen können. Ausserdem werden sich die spitalinternen Gegebenheiten von Spital zu Spital unterscheiden, so dass für jedes Spital womöglich ein eigenes optimales Vergütungssystem zu bestimmen wäre. Darüber hinaus verlangt die optimale Vergütung die Beobachtung und Verifizierung der Behandlungskosten für jeden Patienten. Auch dieser Vorgang wird mit einigen Kosten verbunden sein. Es ist deshalb gut vorstellbar, dass die Kosten der Umsetzung des optimalen Vergütungssystems so hoch sind, dass ein anderes Vergütungssystem, welches zwar nicht die second best-Allokation erreicht, jedoch zu viel geringeren Kosten umzusetzen ist, im Endeffekt vorteilhafter ist als das „optimale“ System.

Doch selbst wenn dies der Fall wäre, bleiben für den Auftraggeber die Erkenntnisse der Analyse von Bedeutung. Denn um entscheiden zu können, ob sich die Umsetzung des optimalen Systems lohnt oder nicht, muss der Auftraggeber zumindest eine Abschätzung des Nutzengewinns vornehmen. Hierzu sind die Ergebnisse der Analyse notwendig.

Zum Beispiel gibt es Schätzungen aus den USA, dass bei einigen DRGs des Medicare

Vergütungssystems bis zu 60% der Behandlungskosten hätten eingespart werden könnten, wenn anstatt der DRG-Fallpauschale ein „optimales“ Vergütungssystem, welches ähnlich wie jenes in Kapitel 8 auf einen wahrheitsgemässen Bericht der Fallschwere durch das Spital abszielt, zum Einsatz gekommen wäre [Chalkley/Malcomson (2002)]. Angesichts dieses Einsparpotentials wäre eine Umsetzung des optimalen Systems sicherlich zu überlegen. Den Einschätzungen der möglichen Einsparung liegt allerdings die Annahme zugrunde, dass die jeweiligen Spitäler lediglich ihren Gewinn maximieren, ohne ein eigenes Interesse an der Behandlung bzw. der Behandlungsqualität zu haben. Im Modell mit doppeltem Unterstellungsverhältnis resultiert eine solche Zielfunktion in Kapitel 7, wenn zwischen Management und Arzt ein PA-Verhältnis vorliegt, die freien Mittel für den Arzt aber exogen gegeben und vom Management nicht zu beeinflussen sind. Bei diesem Mechanismus werden die Präferenzen des Arztes hinsichtlich der Behandlungsqualität vom Management nicht berücksichtigt, und das Management verfolgt nur das Ziel, die Gewinnfunktion

$$EG = \int_{\underline{\theta}}^{\bar{\theta}} (k(\theta) - r(\theta) (p_1 q_1^*(c^d(\theta), p_1, \cdot) + p_2 q_2^*(c^d(\theta), p_1, \cdot)) - \bar{\varphi}(p_1)) f(\theta) d\theta \quad (239)$$

mit der Bestimmung des Preises  $p_1$  und der Budgetierung  $c^d(\theta)$  zu maximieren.

Angenommen, der Entscheidungsprozess ist aber jener wie in Abschnitt 7.6, bei dem Management und Arzt die freien Mittel und die Budgetierung verhandeln. Die dortige Analyse zeigt, dass bei diesem Entscheidungsprozess im Modell mit doppeltem Unterstellungsverhältnis zum einen die Präferenzen des Arztes hinsichtlich der Behandlung aus Sicht des Auftraggebers in der Zielfunktion eine Rolle spielen und zum anderen die Bereitschaft des Managements, Massnahmen zur Verbesserung der Kosteneffizienz durchzuführen umso geringer ist, je weniger das dafür anfallende Arbeitsleid in den Verhandlungen berücksichtigt wird und je grösser die Verhandlungsmacht des Arztes ist. Ob die Grössenordnungen der geschätzten Einsparungen auch bei diesem Entscheidungsprozess resultieren ist fraglich.

### 9.3 Beitrag zur Theorie der Spitalvergütung und Ausblick

Die aus der Analyse in dieser Arbeit gewonnenen Ergebnisse sind für das hier entwickelte Modell mit doppeltem Unterstellungsverhältnis mit den getroffenen Annahmen gültig. So bringt es die Art der Modellierung mit sich, dass konkrete Entscheidungsprozesse angenommen werden müssen, was die Allgemeingültigkeit entsprechend einschränkt. Zwar sind die hier angenommenen Prozesse für ein Spital plausibel, doch sind auch andere Prozesse denkbar, die eventuell zu anderen Ergebnissen führen. Zum Beispiel sind beim internen Entscheidungsmechanismus auch Mischformen vorstellbar, bei denen nur die freien Mittel verhandelt werden, die Budgetierung aber vom Management autonom festgelegt wird.

Oder die Verhandlungen werden nicht über eine Nash-Verhandlungslösung dargestellt, so dass generell aus den Verhandlungen zwischen Arzt und Management keine Pareto-effiziente Allokation resultiert. Zudem verlangt die an sich recht komplexe Modellstruktur vereinfachende Annahmen, die jedoch, wie die angenommene Risikoneutralität, auch anderswo getroffen werden, die allgemeine Aussagekraft der Ergebnisse jedoch weiter verringern.

Der wesentliche Beitrag dieser Arbeit besteht darin, dass die Problematik der spitalinternen Prozesse beleuchtet und in der Modellierung umfassend umgesetzt wurde. Hierarchische Modelle sind in der Theorie der Spitalvergütung zwar nicht neu, jedoch finden sich in der Literatur nur sehr wenige Beiträge, in denen die spitalinterne Interaktion zwischen Management und Arzt explizit formuliert wird. Neure Beispiele sind die Artikel von Jelovac/Macho-Stadler (2002) oder Boadway et al. (2004). Jelovac und Macho-Stadler vergleichen in ihrem Modell einen spitalinternen PA-Mechanismus mit einer organisatorischen Struktur, bei der der Auftraggeber direkt mit dem Arzt einen Vertrag abschliesst. Hinsichtlich der asymmetrischen Informationsverteilung entspricht das Modell von Jelovac und Macho-Stadler der Situation in Kapitel 7. Das Modell von Boadway et al. betrachtet eine Situation mit asymmetrischer Information hinsichtlich der Fallschwere, allerdings werden nicht mehrere organisatorische Strukturen verglichen. Im Vergleich zu diesen beiden Beiträgen ist die vorliegende Arbeit weitergehend: Zum einen werden verschiedene Szenarien der asymmetrischen Informationsverteilung in einem Modell betrachtet. Zum anderen wird mit der NV-Lösung auch ein spitalinterner Mechanismus analysiert, der sich konzeptionell vom üblichen PA-Mechanismus unterscheidet. In vergleichbarer Weise wird in keiner der bisher bestehenden Arbeiten auf dieses Thema eingegangen.

Ausserdem bietet das hier entwickelte Modell eine ganze Reihe von Anknüpfungspunkten für zukünftige Forschungsarbeiten. Zunächst sind hier Modellerweiterungen zu nennen, welche allgemeingültigere Aussagen zulassen, bzw. den Einfluss von Aspekten abdecken, die hier weggelassen wurden. Zu erwähnen wären zum Beispiel die Annahme von Risikoaversion auch beim Management, Einschränkungen bei den freien Mitteln in Form von einer bindenden Mindestbedingung sowie die Betrachtung von multiplen Gleichgewichten oder von Randlösungen, die bei den mehrfachen Maximierungsproblemen natürlich immer auftreten können, hier aber per Annahmen ausgeschlossen wurden. Sicher wäre es auch interessant, den jeweiligen Unterschied zwischen der PA-Lösung und der NV-Lösung zu quantifizieren, um bereits im Rahmen des Modells Aussagen darüber treffen zu können, ob und wenn ja bei welchen Parameterkonstellationen Unterschiede resultieren, die tatsächlich ein relevantes Ausmass haben.

Eine konsequente Weiterentwicklung des Modells bestünde aber vor allem darin, spital-



intern mehr als nur einen Arzt bzw. eine Abteilung zu betrachten, um die Interaktionen zwischen den verschiedenen klinischen Bereichen, die ein wesentliches Merkmal der medizinischen Leistungserbringung in einem Spital sind, in die Modellierung der spitalinternen Entscheidungsprozesse aufnehmen zu können. In einem realen Spital nimmt eine klinische Abteilung für die Behandlung eines Patienten in sehr grossem Umfang die Leistungen anderer Abteilungen oder Funktionsbereiche in Anspruch. Am Universitätsspital Zürich zum Beispiel entspricht bei vielen Abteilungen das Volumen der mit Preisen bewerteten Leistungen, die von anderen Bereichen erbracht wurden, dem Volumen der Kosten, die der Abteilung direkt zugeschrieben werden. In erster Linie sind dies Leistungen von diagnostischen Abteilungen wie Röntgen, Konsiliarleistungen von Medizinern andere Abteilungen oder Leistungen des OP-Bereichs.

Aus der Inanspruchnahme von Leistungen anderer Bereiche durch die behandelnde Abteilung entsteht im Spital eine Trennung von Kostenverursachung und Kostenverantwortung, da die Entscheidung, ob eine Leistung angefordert werden soll, zu einem grossen Teil vom Arzt der behandelnden Abteilung getroffen wird, für die Kosten der Leistung ist jedoch zunächst die erbringende Abteilung verantwortlich. Zum Beispiel entscheidet bei einer Operation grundsätzlich der Anästhesist über die Anästhesieleistungen. Der Arzt der behandelnden Abteilung entscheidet jedoch über die Art der Operation und damit über das während der OP notwendige medizinische Personal und die dafür notwendigen medizinischen Materialien. Mit den dafür anfallenden Kosten wird jedoch der OP-Bereich belastet.

Im internen Rechnungswesen eines Betriebes werden die innerbetrieblichen Leistungsflüsse zwischen den Abteilungen über die so genannte „innerbetriebliche Leistungsverrechnung“ abgebildet. Meist werden die von einer Abteilung in Anspruch genommenen Leistungsmengen anderer Bereiche mit Verrechnungspreisen bewertet, um das Volumen der Inanspruchnahme in Geldeinheiten zu ermitteln. Hinsichtlich des spitalinternen Entscheidungsprozesses ist nun eine wichtige Frage, wie mit diesen Informationen umgegangen wird. Entscheidend ist vor allem, ob die Informationen dazu verwendet werden, Kostenverursachung und Kostenverantwortung wieder zusammenzubringen, indem die in der innerbetrieblichen Leistungsverrechnung ermittelten Kosten der in Anspruch genommenen Leistungen der anfordernden Abteilung belastet werden.

Wenn im Rahmen des Modells mit doppeltem Unterstellungsverhältnis spitalintern mehrere Ärzte bzw. Abteilungen angenommen werden, die gegenseitig Leistungen in Anspruch nehmen, bestünde also ein weiteres Unterscheidungsmerkmal zweier Entscheidungsprozesse bei der Berücksichtigung der innerbetrieblichen Leistungsverrechnung in der Fallbudgetierung  $c^d(\theta)$ . Ist die innerbetriebliche Leistungsverrechnung nicht Bestandteil der Budgetierung, kann der behandelnde Arzt mit der Anforderung von Leistungen einer anderen

Abteilung seinen Nutzen maximieren, ohne dabei die Budgetbeschränkung  $c^d(\theta)$  beachten zu müssen. Eine Budgetkürzung wird sich in diesem Fall eher in einer Verringerung der Abteilungsleistungen und in einer Ausweitung der Anforderungen der Leistungen anderer Abteilungen auswirken. Ist die innerbetriebliche Leistungsverrechnung dagegen Bestandteil der Budgetierung und kennt der Arzt die Verrechnungspreise der Leistungen, dann bestimmt der Arzt auch die angeforderte Menge entsprechend dem - um die angeforderte Menge erweiterten - Maximierungsproblem bei Budgetbeschränkung in Abschnitt 6.3.2 in Abhängigkeit der Preise aller Leistungen und der Budgetbeschränkung.

Bei gegebener Budgetierung resultieren bei beiden Entscheidungsprozessen unterschiedliche zur Behandlung eingesetzte Mengen der verschiedenen Leistungen. Hinzu kommt, dass auch die Bestimmung der Leistungsmengen durch den erbringenden Arzt, der seinerseits ebenfalls Budgetvorgaben zu beachten hat, in Abhängigkeit von den verschiedenen Mechanismen steht. Wenn also spitalintern eine Abteilung zur Behandlung ihrer Patienten die Leistungen anderer Abteilungen in Anspruch nimmt, dann ist die Information, ob und wie die innerbetriebliche Leistungsverrechnung in einem Spital etabliert ist, für den Auftraggeber elementar bei der Bestimmung des optimalen Vergütungssystems.

## A Anhang

### A.1 Unabhängigkeit der Mengen $q_1$ und $q_2$ von der Fallschwere

Sei

$$|H| = \begin{vmatrix} 0 & -p_1 & -p_2 \\ -p_1 & \theta V_{q_1 q_1}^* & \theta V_{q_1 q_2}^* \\ -p_2 & \theta V_{q_2 q_1}^* & \theta V_{q_2 q_2}^* \end{vmatrix} \quad (240)$$

die Determinante der Hesse-Matrix des Maximierungsproblems des Arztes mit den Bedingungen erster Ordnung (48) - (50) im Optimum (die Argumente und Klammern sind zur übersichtlicheren Darstellung unterdrückt). Sei weiter

$$|H_1| = \begin{vmatrix} 0 & -p_1 & -p_2 \\ -V_{q_1}^* & \theta V_{q_1 q_1}^* & \theta V_{q_1 q_2}^* \\ -V_{q_2}^* & \theta V_{q_2 q_1}^* & \theta V_{q_2 q_2}^* \end{vmatrix} \quad (241)$$

$$|H_2| = \begin{vmatrix} 0 & 0 & -p_2 \\ -p_1 & -V_{q_1}^* & \theta V_{q_1 q_2}^* \\ -p_2 & -V_{q_2}^* & \theta V_{q_2 q_2}^* \end{vmatrix} \quad (242)$$

$$|H_3| = \begin{vmatrix} 0 & -p_1 & 0 \\ -p_1 & \theta V_{q_1 q_1}^* & -V_{q_1}^* \\ -p_2 & \theta V_{q_2 q_1}^* & -V_{q_2}^* \end{vmatrix} \quad (243)$$

Für  $|H_2|$  und  $|H_3|$  gilt mit (57)

$$|H_2| = -p_2(p_1 V_{q_2}^* - p_2 V_{q_1}^*) = 0 \quad (244)$$

und

$$|H_3| = -p_1(p_2 V_{q_1}^* - p_1 V_{q_2}^*) = 0. \quad (245)$$

Mit der Cramer'schen Regel folgt

$$\frac{\partial q_1^*(\cdot)}{\partial \theta} = \frac{|H_2|}{|H|} = 0, \quad \frac{\partial q_2^*(\cdot)}{\partial \theta} = \frac{|H_3|}{|H|} = 0 \quad (246)$$

## A.2 Einfluss der Verhandlungsmacht und der Vergütungsparameter bei Risikoaversion des Arztes

Im Folgenden seien  $\pi_{tt} = \frac{\partial \pi_t}{\partial t^d}$ ,  $\pi_{cc} = \frac{\partial \pi_c}{\partial c}$ ,  $\pi_{tw} = \frac{\partial \pi_t}{\partial w}$  usw. die Ableitungen der Zielfunktion (106)

$$\pi(\theta) = G(\theta)^w (U(\theta) - u)^{1-w} = (k(\theta) - r(\theta)c^d(\theta) - t^d(\theta))^w (H(t^d(\theta)) + \theta V^*(.) - u)^{1-w} \quad (247)$$

wobei die Fallschwere und die Indizes zur vereinfachten Darstellung unterdrückt werden.

### A.2.1 Auswirkung der Verhandlungsmacht $w$ auf die freien Mittel $t^d(\theta)$ und die Budgetierung $c^d(\theta)$

Es sei

$$|H| = \begin{vmatrix} \pi_{tt} & \pi_{ct} \\ \pi_{tc} & \pi_{cc} \end{vmatrix} > 0 \quad (248)$$

die Determinante der Hesse - Matrix der Zielfunktion (247). Im Maximum ist sie notwendigerweise positiv. Ausserdem sei

$$|H_{1w}| = \begin{vmatrix} -\pi_{tw} & \pi_{tc} \\ -\pi_{cw} & \pi_{cc} \end{vmatrix} \quad \text{und} \quad |H_{2w}| = \begin{vmatrix} \pi_{tt} & -\pi_{tw} \\ \pi_{ct} & -\pi_{cw} \end{vmatrix}.$$

Mit der Cramer'schen Regel gilt

$$\frac{\partial t}{\partial w} = \frac{|H_{1w}|}{|H|} \quad (249)$$

und

$$\frac{\partial c}{\partial w} = \frac{|H_{2w}|}{|H|}. \quad (250)$$

Für die zweiten Ableitungen bzw. für die Kreuzableitung von (247) ergibt sich

$$\begin{aligned} \pi_{tt} &= -w(1-w) [G^{w-2}U^{1-w} + 2G^{w-1}U^{-w}H'(t) + G^wU^{-w-1}H'(t)^2] + \\ &+ (1-w) \left(\frac{G}{U}\right)^w H''(t) < 0, \end{aligned} \quad (251)$$

und

$$\begin{aligned} \pi_{cc} = & -w(1-w) \left[ r^2 G^{w-2} U^{1-w} + 2r \lambda^*(.) G^{w-1} U^{-w} H'(t) + G^w U^{-w-1} \lambda^*(.)^2 \right] + \\ & + (1-w) \left( \frac{G}{U} \right)^w \frac{\partial \lambda^*(.)}{\partial c} < 0 \end{aligned} \quad (252)$$

sowie

$$\begin{aligned} \pi_{tc} = & -w(1-w) \left[ r G^{w-2} U^{1-w} + \lambda^*(.) G^{w-1} U^{-w} \right] - \\ & -w(1-w) \left[ r H'(t) G^{w-1} U^{-w} + G^w U^{-w-1} \lambda^*(.) H'(t) \right] < 0. \end{aligned} \quad (253)$$

Mit (107) folgt

$$\pi_{tc} = r(\pi_{tt} - (1-w) \left( \frac{G}{U} \right)^w H''(t)) \quad (254)$$

und

$$\pi_{cc} = r\pi_{tc} + (1-w) \left( \frac{G}{U} \right)^w \frac{\partial \lambda^*(.)}{\partial c}. \quad (255)$$

Ableiten der Bedingungen erster Ordnung (107) und (108) nach  $w$  ergibt für  $\pi_{tw}$  bzw.  $\pi_{cw}$  im Optimum

$$\begin{aligned} \pi_{tw} = & - \left( \frac{U}{G} \right)^{1-w} - H'(t) \left( \frac{G}{U} \right)^w + \\ & + \ln \left( \frac{G}{U} \right) \underbrace{\left[ (1-w) \left( \frac{G}{U} \right)^w H'(t) - w \left( \frac{U}{G} \right)^{1-w} \right]}_{=0} < 0 \end{aligned} \quad (256)$$

bzw.

$$\begin{aligned} \pi_{cw} = & -r \left( \frac{U}{G} \right)^{1-w} - \lambda^*(.) H'(t) \left( \frac{G}{U} \right)^w + \\ & + \ln \left( \frac{G}{U} \right) \underbrace{\left[ \lambda^*(.) (1-w) \left( \frac{G}{U} \right)^w H'(t) - r w \left( \frac{U}{G} \right)^{1-w} \right]}_{=0} < 0, \end{aligned} \quad (257)$$

wobei der jeweilige Ausdruck in den eckigen Klammern wegen (107) im Maximum 0 ist. Mit (108) gilt deswegen

$$\pi_{cw} = r\pi_{tw}. \quad (258)$$

Mit diesen Ergebnissen folgt für  $|H_{1w}|$

$$|H_{1w}| = -\pi_{tw} \left[ r\pi_{tc} + (1-w) \left( \frac{G}{U} \right)^w \frac{\partial \lambda^*(.)}{\partial c} \right] + r\pi_{tw}\pi_{tc} = -\pi_{tw}(1-w) \left( \frac{G}{U} \right)^w \frac{\partial \lambda^*(.)}{\partial c} < 0, \quad (259)$$

und daher

$$\frac{\partial t}{\partial w} = \frac{|H_{1w}|}{|H|} < 0. \quad (260)$$

Für  $|H_{2w}|$  hingegen folgt

$$|H_{2w}| = -\pi_{tt}r\pi_{tw} + \left[ r\pi_{tt} - r(1-w) \left( \frac{G}{U} \right)^w H''(t) \right] \pi_{tw} = -\pi_{tw}r(1-w) \left( \frac{G}{U} \right)^w H''(t) < 0 \quad (261)$$

und somit

$$\frac{\partial c}{\partial w} = \frac{|H_{2w}|}{|H|} < 0. \quad (262)$$

### A.2.2 Auswirkung der Kostenbeteiligung $r(\theta)$ und der Pauschale $k(\theta)$ auf die freien Mittel $t^d(\theta)$ und die Budgetierung $c^d(\theta)$

Ableiten der Bedingungen erster Ordnung (107) und (108) nach  $r$  ergibt

$$\pi_{tr} = -w(1-w)c[G^{w-2}U^{1-w} + G^{w-1}U^{-w}H'(t)] < 0 \quad (263)$$

und

$$\pi_{cr} = -w(1-w)c[G^{w-2}U^{1-w}r + G^{w-1}U^{-w}\lambda^*(.)] - w \left( \frac{U}{G} \right)^{1-w} < 0, \quad (264)$$

und damit

$$\pi_{cr} = r\pi_{tr} - w \left( \frac{U}{G} \right)^{1-w}. \quad (265)$$

Ableiten der Bedingungen erster Ordnung (107) und (108) nach  $k$  ergibt dagegen

$$\pi_{tk} = w(1-w)[G^{w-2}U^{1-w} + G^{w-1}U^{-w}H'(t)] > 0 \quad (266)$$

und

$$\pi_{ck} = w(1-w)[G^{w-2}U^{1-w}r + G^{w-1}U^{-w}\lambda^*(.)] > 0, \quad (267)$$

und damit

$$\pi_{ck} = r\pi_{tk}. \quad (268)$$

Sei

$$|H_{1r}| = \begin{vmatrix} -\pi_{tr} & \pi_{tc} \\ -\pi_{cr} & \pi_{cc} \end{vmatrix} \quad \text{und} \quad |H_{2r}| = \begin{vmatrix} \pi_{tt} & -\pi_{tr} \\ \pi_{ct} & -\pi_{cr} \end{vmatrix}.$$

bzw.

$$|H_{1k}| = \begin{vmatrix} -\pi_{tk} & \pi_{tc} \\ -\pi_{ck} & \pi_{cc} \end{vmatrix} \quad \text{und} \quad |H_{2k}| = \begin{vmatrix} \pi_{tt} & -\pi_{tk} \\ \pi_{ct} & -\pi_{ck} \end{vmatrix}.$$

Es ergibt sich

$$\begin{aligned} |H_{1r}| &= -\pi_{tr}\pi_{cc} + \pi_{cr}\pi_{tc} = -\pi_{tr} \left[ r\pi_{tc} + (1-w) \left( \frac{G}{U} \right)^w \frac{\partial \lambda^*(.)}{\partial c} \right] + \left[ r\pi_{tr} - w \left( \frac{U}{G} \right)^{1-w} \right] \pi_{tc} \\ &= -\pi_{tr}(1-w) \left( \frac{G}{U} \right)^w \frac{\partial \lambda^*(.)}{\partial c} - w\pi_{tc} \left( \frac{U}{G} \right)^{1-w} \pi_{tc} \leq 0 \end{aligned} \quad (269)$$

und

$$\begin{aligned} |H_{2r}| &= -\pi_{tt}\pi_{cr} + \pi_{tc}\pi_{tr} = \pi_{tt} \left[ r\pi_{tr} - w \left( \frac{U}{G} \right)^{1-w} \right] + \left[ r\pi_{tt} - r(1-w) \left( \frac{G}{U} \right)^w H''(t) \right] \pi_{tr} \\ &= \pi_{tt}w \left( \frac{U}{G} \right)^{1-w} - r(1-w) \left( \frac{G}{U} \right)^w H''(t)\pi_{tr} < 0 \end{aligned} \quad (270)$$

bzw.

$$\begin{aligned}
|H_{1k}| &= -\pi_{tk}\pi_{cc} + \pi_{ck}\pi_{tc} = -\pi_{tk} \left[ r\pi_{tc} + (1-w) \left( \frac{G}{U} \right)^w \frac{\partial \lambda^*(.)}{\partial c} \right] + r\pi_{tk}\pi_{tc} \\
&= -\pi_{tk}(1-w) \left( \frac{G}{U} \right)^w \frac{\partial \lambda^*(.)}{\partial c} > 0
\end{aligned} \tag{271}$$

und

$$\begin{aligned}
|H_{2k}| &= -\pi_{tt}\pi_{ck} + \pi_{tc}\pi_{tk} = \pi_{tt}r\pi_{tk} + \left[ r\pi_{tt} - r(1-w) \left( \frac{G}{U} \right)^w H''(t) \right] \pi_{tk} \\
&= -r(1-w) \left( \frac{G}{U} \right)^w H''(t) \pi_{tk} \geq 0
\end{aligned} \tag{272}$$

Mit der Cramer'schen Regel gilt deshalb

$$\frac{\partial t}{\partial r} = \frac{|H_{1r}|}{|H|} \leq \geq 0 \quad \text{und} \quad \frac{\partial c}{\partial r} = \frac{|H_{2r}|}{|H|} < 0 \tag{273}$$

bzw.

$$\frac{\partial t}{\partial k} = \frac{|H_{1k}|}{|H|} > 0 \quad \text{und} \quad \frac{\partial c}{\partial k} = \frac{|H_{2k}|}{|H|} > 0. \tag{274}$$



### A.3 Das Arbeitsleid $\varphi$ als direkte Funktion des Preises $p_1$

Es wird gezeigt, dass das Arbeitsleid als direkte Funktion des Preises  $p_1$  fallend und konkav ist.

Da die Funktion  $p_1(e)$  streng monoton fallend und konvex ist, ist auch die Umkehrfunktion

$$p_1^{-1} \equiv e(p_1) \quad (275)$$

streng monoton fallend und konvex. Also gilt

$$\frac{\partial e(p_1)}{\partial p_1} < 0 \quad (276)$$

sowie

$$\frac{\partial^2 e(p_1)}{(\partial p_1)^2} > 0. \quad (277)$$

Damit gilt für das Arbeitsleid

$$\frac{\partial \varphi(e(p_1))}{\partial p_1} = \varphi'(e(p_1)) \frac{\partial e(p_1)}{\partial p_1} < 0 \quad (278)$$

sowie

$$\frac{\partial^2 \varphi(e(p_1))}{(\partial p_1)^2} = \varphi''(e(p_1)) \left[ \frac{\partial e(p_1)}{\partial p_1} \right]^2 + \varphi'(e(p_1)) \frac{\partial^2 e(p_1)}{(\partial p_1)^2} > 0. \quad (279)$$

Wegen den angenommenen Inada-Bedingungen gilt ausserdem

$$e(p_1^{max}) = 0 \quad (280)$$

und

$$\lim_{p_1 \rightarrow 0} e(p_1) \rightarrow \infty. \quad (281)$$

sodass

$$\varphi(p_1^{max}) = 0, \quad \lim_{p \rightarrow 0} \varphi(p) \rightarrow \infty. \quad (282)$$

folgt.

## A.4 Globale Bedingung für wahrheitsgemässe Berichterstattung des Managements

Für wahrheitsgemässe Berichterstattung des Managements für alle möglichen Fallschweren ist Bedingung (169)  $\dot{G}(\theta) \equiv \frac{dG(\theta)}{d\theta} = V^*(c^m(\theta)) > 0$  zusammen mit Bedingung (170)  $\dot{c}^m(\theta) \geq 0$  hinreichend. Der Beweis ist aus Laffont/Tirole (1993), S. 121 übernommen.

Es sei  $G(\theta, \hat{\theta})$  der Gewinn des Managements, wenn die tatsächliche Fallschwere des Patienten den Wert  $\theta$  hat, das Management aber  $\hat{\theta}$  angibt. Das Management berichtet die Fallschwere wahrheitsgemäss für jede Fallschwere, wenn

$$G(\theta, \theta) \geq G(\theta, \hat{\theta}) \quad \forall \theta, \hat{\theta} \in \Theta. \quad (283)$$

Es  $G_1(\theta, \hat{\theta})$  die Ableitung nach dem ersten Argument,  $G_2(\theta, \hat{\theta})$  die Ableitung nach dem zweiten und  $G_{12}(\theta, \hat{\theta})$  die Kreuzableitung von  $G(\theta, \hat{\theta})$ . Ausdruck (283) lässt sich schreiben als

$$\int_{\hat{\theta}}^{\theta} G_2(\theta, x) dx \geq 0. \quad (284)$$

Wenn wahrheitsgemässe Berichterstattung optimal für den Arzt ist, dann muss  $G_2(\theta, \hat{\theta})|_{\hat{\theta}=\theta} = 0$  gelten. Damit folgt

$$\begin{aligned} \int_{\hat{\theta}}^{\theta} G_2(\theta, x) &= \int_{\hat{\theta}}^{\theta} G_2(\theta, x) - G_2(x, x) dx \\ &= \int_{\hat{\theta}}^{\theta} \int_x^{\theta} G_{21}(y, x) dy dx \geq 0. \end{aligned} \quad (285)$$

Mit (168) gilt  $G_{21}(\theta, \theta) = \frac{\partial \lambda^*(\theta, \cdot)}{\partial \theta} \dot{c}^m(\theta)$ , wobei  $\frac{\partial \lambda^*(\theta, \cdot)}{\partial \theta} > 0$  (Ausdruck (61)). Wenn  $\hat{\theta} < \theta$ , dann  $x < \theta$  und (285) ist erfüllt, wenn  $\dot{c}^m(\theta) \geq 0$ . Wenn dagegen  $\hat{\theta} > \theta$  ist, dann  $x > \theta$  und (285) ist ebenfalls erfüllt, wenn die Budgetierung mit der Fallschwere steigt.

## A.5 Globale Bedingung für wahrheitsgemässe Berichterstattung des Managements bei spitalinterner asymmetrischer Information

Es sei  $\tilde{G}(\theta, \hat{\theta}^m) = t^m(\hat{\theta}^m) + \theta V^*(c^m(\hat{\theta}^m)) - h(\theta)V^*(c^m(\hat{\theta}^m))$  der virtuelle Gewinn des Managements (199). Für die Kreuzableitung von (199) gilt

$$\tilde{G}_{21}(\theta, \hat{\theta}) = (1 - h'(\theta)) \frac{\partial \lambda^*(\theta, \cdot)}{\partial \theta} c^m(\theta). \quad (286)$$

Da  $(1 - h'(\theta)) > 0$  und  $\frac{\partial \lambda^*(\theta, \cdot)}{\partial \theta} > 0$  ist der Ausdruck (286) ebenfalls grösser null und der Beweis erfolgt äquivalent zu Anhang A.4.

## A.6 Nash-Verhandlungslösung bei spitalinterner asymmetrischer Information hinsichtlich der Fallschwere

Wir nehmen an, dass, ebenso wie das Management bei der PA-Lösung bei spitalinterner asymmetrischer Information hinsichtlich der Fallschwere den erwarteten Gewinn maximiert, Management und Arzt im Rahmen der Verhandlungen die Maximierung ihres Anteils am erwarteten gemeinsamen Nutzenüberschuss zum Ziel haben. Unter Berücksichtigung der Bedingung für die wahrheitsgemässe Berichterstattung durch den Arzt, sowie der Teilnahmebedingung des Arztes und der internen Teilnahmebedingung des Managements für jede Fallschwere bestimmt die NV-Lösung also die optimalen Pfade der Berichtsstrategie  $\hat{\theta}^{m*}(\theta)$  und der freien Mittel  $t^{d*}(\theta)$  zur Maximierung des folgenden Problems:

$$\max_{\hat{\theta}^m(\theta), t^d(\theta)} \left( E[t^m(\hat{\theta}^m(\theta)) - t^d(\theta)] \right)^{1-w} \left( E[t^d(\theta) + \theta V^*(c^m(\hat{\theta}^m(\theta))) - u] \right)^{1-w} \quad (287)$$

unter den Bedingungen

$$\dot{U}(\theta) = V^*(c^m(\hat{\theta}^m(\theta))) > 0 \quad \forall \theta \in \Theta \quad (288)$$

$$U(\underline{\theta}) \geq u \quad (289)$$

$$G(\theta) \geq 0 \quad \forall \theta \in \Theta. \quad (290)$$

Wegen der Bedingung für wahrheitsgemässe Berichterstattung (288) wird der ex-post Nutzen des Arztes  $U(\theta) \quad \forall \theta \in \Theta$  bei gegebenem Vergütungssystem aber eindeutig durch den verhandelten Pfad der Berichtstrategie  $\hat{\theta}^{m*}(\theta)$  und einer Pauschale  $\epsilon$  mit

$$U(\theta) = \int_{\underline{\theta}}^{\theta} V^*(c^m(\hat{\theta}^{m*}(\tilde{\theta}))) d\tilde{\theta} + u + \epsilon \quad (291)$$

definiert. Mit  $\tilde{U}(\theta) = U(\theta) - \epsilon$  lässt sich das Maximierungsproblem auch äquivalent formulieren als

$$\max_{\epsilon, \hat{\theta}^m(\theta)} \left( E[t^m(\hat{\theta}^m(\theta)) + \theta V^*(c^m(\hat{\theta}^m(\theta))) - E[\tilde{U}(\hat{\theta}^m(\theta)) - \epsilon] \right)^w \left( E[\tilde{U}(\hat{\theta}^m(\theta))] + \epsilon - u \right)^{1-w} \quad (292)$$

s.t.

$$\dot{U}(\theta) = \dot{\tilde{U}}(\theta) = V^*(c^m(\hat{\theta}^{m*}(\theta))) \quad (293)$$

$$\epsilon \geq 0 \quad (294)$$

$$G(\theta) \geq 0 \quad \forall \theta \in \Theta. \quad (295)$$

Bei dieser Formulierung verhandeln Arzt und Management mit  $\epsilon^*$  den Betrag, mit dem der Nutzen des Arztes bei der Behandlung eines Patienten der niedrigsten Fallschwere  $\underline{\theta}$  über dem Reservationsnutzen  $u$  liegt. Mit dem ebenfalls in der Verhandlung bestimmten Pfad der Berichtsstrategie  $\hat{\theta}^m(\theta)$  wird damit wegen der Anreizbedingung (293) der ex-post Nutzen  $U(\theta)$  des Arztes und damit auch der erwartete Nutzen bestimmt.

Wenn die Bedingungen (294) und (295) nicht binden, wird die Pauschale  $\epsilon$  so gewählt, dass für den Arzt ein Anteil am erwarteten gemeinsamen Nutzenüberschuss resultiert, der seiner Verhandlungsmacht entspricht, wobei wie bereits erwähnt die durch die Bedingung (293) implizierte Informationsrente in den Anteil für den Arzt einberechnet wird. Das bedeutet, die Informationsrente für den Arzt wird vollständig zwischen Arzt und Management internalisiert und die Verhandlungslösung maximiert den gemeinsamen erwarteten Nutzenüberschuss von Arzt und Management. Der optimale Pfad der Berichterstattung maximiert also ebenfalls das Problem

$$\max_{\hat{\theta}^m(\theta)} E[N\dot{U}(\theta)] = \int_{\underline{\theta}}^{\bar{\theta}} [t^m(\hat{\theta}^m(\theta)) + \theta V^*(c^m(\hat{\theta}^m(\theta)))] f\theta d\theta - u \quad (296)$$

unabhängig von der Aufteilung des erwarteten Nutzenüberschusses zwischen Arzt und Management, d.h. unabhängig von der Höhe des Zuschlages  $\epsilon$ . Es ist offensichtlich, dass dieses Maximierungsproblem mit jenem des Managements bei vollständiger spitalinterner Information hinsichtlich der Fallschwere (168) identisch ist. Wenn die Bedingungen (294) und (295) nicht binden, führt also das Vergütungssystem, das bei vollständiger spitalinternen Information optimal ist und mit den Gleichungen (181) und (182) beschrieben wird, auch hier zu wahrheitsgemässer Berichterstattung durch das Spital, ungeachtet der tatsächlichen spitalinternen Informationsasymmetrie.

Wir zeigen im Folgenden, dass die Bedingungen (294) und (295) unter dem bei vollständiger spitalinterner Information optimalen und mit (181) und (182) beschriebenen Vergütungssystem tatsächlich nicht binden, wenn der Arzt gegenüber den Management über eine gewisse Verhandlungsmacht verfügt.

Hierzu stellen wir fest, dass es unter dem optimalen Vergütungssystem bei vollständiger spitalinterner Information hinsichtlich der Fallschwere genau ein  $\bar{w} \in (0, 1)$  existiert, für das gilt:

$$\epsilon^* > 0 \text{ und } G(\theta) > 0 \quad \forall \theta \in \Theta, \text{ wenn } w < \bar{w}. \quad (297)$$

Zum Beweis führen wir an, dass beim unbeschränktes Maximum der Funktion (292) als Bedingung erster Ordnung für das optimale  $\epsilon^*$

$$w \left( E[\tilde{U}(\theta)] + \epsilon^* - u \right) = (1 - w)E[G(\theta)] \quad (298)$$

resultiert. Daraus folgt

$$d\epsilon^*/dw < 0 \quad (299)$$

sowie

$$E[\tilde{U}(\theta)] + \epsilon^* = E[U(\theta)] > u \text{ und } \lim_{w \rightarrow 1} E[U(\theta)] = u. \quad (300)$$

Da annahmegemäss unter der Prämisse der wahrheitsgemässen Berichterstattung (293) verhandelt wird, folgt aber auch  $\dot{U}(\theta) = V^*(c^m(\hat{\theta}^m(\theta))) > 0$  und äquivalent zu (178):

$$E[U(\theta)] - u = \int_{\underline{\theta}}^{\bar{\theta}} h(\theta) V^*(c^m(\hat{\theta}^m(\theta))) f(\theta) d\theta + \epsilon^*. \quad (301)$$

Da  $\int_{\underline{\theta}}^{\bar{\theta}} h(\theta) V^*(c^m(\hat{\theta}^m(\theta))) f(\theta) d\theta > 0$  muss es wegen (299), (300) und (301) genau ein  $\bar{w} < 1$  geben, für das im unbeschränkten Maximum

$$\epsilon^* \begin{cases} > 0 \text{ wenn } w < \bar{w}, \\ = 0 \text{ wenn } w = \bar{w}, \\ < 0 \text{ wenn } w > \bar{w}. \end{cases} \quad (302)$$

gilt.

Auf der anderen Seite gilt  $\bar{w} > 0$ . Aus (298) folgt nämlich auch

$$E[U(\theta)] < E[N\ddot{U}(\theta)] \text{ und } \lim_{w \rightarrow 0} E[U(\theta)] = E[N\ddot{U}(\theta)]. \quad (303)$$

Mit dem optimalen Vergütungssystem bei vollständiger spitalinterner Information hinsichtlich der Fallschwere ist wegen (178) der gemeinsame erwartete Nutzenüberschuss gegeben mit  $E[N\ddot{U}(\theta)] = \int_{\underline{\theta}}^{\bar{\theta}} h(\theta) V^*(c^m(\hat{\theta}^m(\theta))) f(\theta) d\theta + g$ . Mit (303) folgt daher

$$\epsilon^* < g \text{ und } \lim_{w \rightarrow 0} \epsilon^* = g > 0 \quad (304)$$

und damit  $\bar{w} > 0$ . Für  $w \in (0, \bar{w})$  bindet also die Bedingung (294) nicht.

Für  $w \in (0, \bar{w})$  ist darüber hinaus auch die interne Teilnahmebedingung des Managements (295) nicht einschränkend. Da  $G(\theta) = N\ddot{U}(\theta) - U(\theta) + u$  folgt mit  $N\ddot{U}(\theta) = \int_{\underline{\theta}}^{\theta} V^*(c^m(\hat{\theta}^m(\theta))) + g$  und  $U(\theta) = \int_{\underline{\theta}}^{\theta} V^*(c^m(\hat{\theta}^m(\theta))) + u + \epsilon$

$$G(\theta) = g - \epsilon^* > 0 \quad \forall \theta \in \Theta. \quad (305)$$

Wenn der Auftraggeber also das Vergütungssystem einsetzt, das bei vollständiger spitalinterner Information hinsichtlich der Fallschwere optimal ist und wenn die Verhandlungsmacht zwischen Arzt und Management derart verteilt ist, dass  $w \in (0, \bar{w})$  gilt, dann ist das Maximierungsproblem (292) unbeschränkt und aus Sicht des Auftraggebers äquivalent zu dem Maximierungsproblem (296). Das bedeutet, dieses Vergütungssystem impliziert ungeachtet der spitalinternen Informationsasymmetrie die Verhandlung wahrheitsgemässer Berichterstattung (durch das Management), ist mit spitalinterner wahrheitsgemässer Berichterstattung durch den Arzt vereinbar und führt aus Sicht des Auftraggebers zu derselben Allokation, die bei vollständiger spitalinterner Information resultiert. Bei spitalinterner NV-Lösung wird also die spitalinterne Informationsasymmetrie vollständig zwischen Arzt und Management internalisiert wenn  $w \in (0, \bar{w})$  und gelangt insbesondere nicht zum Nachteil des Auftraggebers.

## A.7 Symbole und Ausdrücke

$q_1, q_2$	Von Behandlungsmassnahme 1 bzw. 2 eingesetzte Menge.
$\theta$	$\in [\underline{\theta}, \bar{\theta}]$ Fallschwere eines Patienten.
$p_1, p_2$	Preis für eine Einheit von Behandlungsmassnahme 1 bzw. 2.
$c$	$= p_1 q_1 + p_2 q_2$ realisierte Behandlungskosten eines Patienten.
$c^d(\theta)$	Budgetvorgabe für den Arzt bei der Behandlung eines Patienten.
$t^d(\theta)$	Zusätzliche finanzielle Mittel für den Arzt bei Behandlung eines Patienten.
$c^m(\theta)$	Budgetvorgabe für das Management für die Behandlung eines Patienten.
$t^m(\theta)$	Vergütung des Spitals für die Behandlung eines Patienten.
$\hat{\theta}^d$	Berichtete Fallschwere durch den Arzt.
$\hat{\theta}^m$	Berichtete Fallschwere durch das Management.
$\theta V(q_1, q_2)$	Bewertungsfunktion des Arztes für die Behandlung eines Patienten.
$U(\theta)$	$= t^d(\theta) + \theta V(q_1, q_2)$ Nutzen des Arztes pro Patient.
$\tilde{u}$	ex-ante Reservationsnutzen des Arztes pro Patient.
$u$	ex-post Reservationsnutzen des Arztes pro Patient.
$\lambda(\theta, c^d(\theta))$	Zahlungsbereitschaft des Arztes für eine marginale Ausweitung der Budgetvorgabe $c^d(\theta)$ für die Behandlung eines Patienten (s. Abschnitt (6.3.2)).
$\bar{\varphi}(p_1)$	Arbeitsleid des Managements pro Patient.
$G(\theta)$	$= t^m(\theta) - t^d(\theta)$ Gewinn des Managements bei der Behandlung eines Patienten. Ggf. abzüglich $\bar{\varphi}(p_1)$ .
$N\ddot{U}(\theta)$	$= G(\theta) + U(\theta) - u$ Nutzenüberschuss von Arzt und Management pro Patient. Ggf. abzüglich $\bar{\varphi}(p_1)$ .
$\tilde{g}$	ex-ante Mindestwert des Nutzenüberschuss pro Patient zur Teilnahme des Managements.
$g$	ex-post Mindestwert des Nutzenüberschuss pro Patient zur Teilnahme des Managements.
$\theta B(q_1, q_2)$	Bewertungsfunktion des Auftraggebers für die Behandlung eines Patienten.
$W(\theta)$	$= \theta B(q_1, q_2) - c - t^m(\theta)$ Nutzen des Auftraggebers bei der Behandlung eines Patienten.
$w$	$\in (0, 1)$ Verhandlungsmacht des Managements bei der Nash-Verhandlungslösung.
$1 - w$	Verhandlungsmacht des Arztes bei der Nash-Verhandlungslösung.
$\alpha$	$\in [0, 1]$ Anteil des Arbeitsleides, den das Management in der Verhandlung mit dem Arzt geltend machen kann.

Tabelle 5: Im Modell verwendete Symbole und Ausdrücke. Bei Symbolen mit Fallschwere  $\theta$  in Klammern bezieht sich der Ausdruck auf die Behandlung eines Patienten der Fallschwere, z.B.  $c^d(\theta)$  Budgetvorgabe für den Arzt bei der Behandlung eines Patienten der Fallschwere  $\theta$ .



## Literatur

- [1] K. J. Arrow. Uncertainty and the welfare economics of medical care. *The American Economic Review*, 53(5):941–973, 1963.
- [2] C. L. Ballard, J. B. Shoven und J. Whalley. General equilibrium computations of the marginal welfare costs of taxes in the united states. *American Economic Review*, 75(1):128 – 138, 03 1985.
- [3] P. P. Barros und X. Martinez-Giralt. Models of negotiation and bargaining in health care. In: *The Elgar companion to health economics*, Seiten 233 – 241. Elgar, 2006.
- [4] D. Benz. *Unternehmerische Spitalführung : Ein integriertes Konzept für öffentliche Spitäler in der Schweiz*. Sankt Gallen : [s.n.], 2007.
- [5] F. Breyer. *Mikroökonomik: Eine Einführung*. Springer, 2004.
- [6] F. Breyer und M. Kolmar. *Grundlagen der Wirtschaftspolitik*. Mohr Siebeck, Tübingen, 2. Aufl., 2005.
- [7] F. Breyer, P. Zweifel und M. Kifmann. *Gesundheitsökonomik*. Springer, Berlin, 5. Aufl., 2005.
- [8] R. Boadway, M. Marchand und M. Sato. An optimal contract approach to hospital financing. *Journal of Health Economics*, (23):85–110, 2004.
- [9] H. Bunzemeier et al. Erlösverteilung unter DRG-Bedingungen am Universitätsklinikum Münster. *Das Krankenhaus*, (10):946 – 961, 2010.
- [10] H.-P. Busch. Das Berufsbild Chefarzt im Wandel. *Das Krankenhaus*, (3), 2011.
- [11] M. Chalkley und J. M. Malcomson. Contracting for health services when patient demand does not reflect quality. *Journal of Health Economics*, (17):1–19, 1998.
- [12] M. Chalkley und J. M. Malcomson. Contracting for health services with unmonitored quality. *The Economic Journal*, (108):1093–1110, 1998.
- [13] M. Chalkley und J. M. Malcomson. Government purchasing of health services. In: A. Culyer und J. Newhouse, *Handbook of Health Economics*, Volume 1, Kapitel 15. Elsevier Science, 2000.
- [14] M. Chalkley und J. M. Malcomson. Cost sharing in health service provision: An empirical assessment of cost saving. *Journal of Public Economics*, (84):219–249, 2002.

- [15] A. C. Chiang. *Fundamental Methods of Mathematical Economics*. McGraw-Hill, 3. edition, 1984.
- [16] L. S. Dafny. How do hospitals respond to price changes? *The American Economic Review*, 95(5):1525–1547, 2005.
- [17] A. Donabedian. Evaluating the quality of medical care. *The Milbank Memorial Fund Quarterly*, 44(3):166 – 203, 1966.
- [18] D. Dranove. The industrial organization of health care markets. In: A. Culyer und J.P.Newhouse, *Handbook of Health Economics*, Volume 1, Kapitel 20. Elsie Science, 2000.
- [19] D. Dranove, D. Kessler, M. McClellan und M. Satterthwaite. Is More Information Better? The Effects of Report Cards on Health Care Providers. *The Journal of Political Economy*, 111(3):555–588, 2003.
- [20] K. Eggleston. Multitasking and mixed systems for provider payment. *Journal of Health Economics*, (24):211–223, 2005.
- [21] R. P. Ellis. Creaming, skimping and dumping: provider competition on the intensive and extensive margins. *Journal of Health Economics*, (17):537–555, 1998.
- [22] R. P. Ellis und T. G. McGuire. Provider behavior under prospective reimbursement: Cost sharing and supply. *Journal of Health Economics*, 5(2):129 – 151, 1986.
- [23] C. Ernst. *Krankenhaus-Controlling und monetäre Anreize für leitende Ärzte : eine Agency - theoretische Analyse*. Gabler Edition Wissenschaft. Deutscher Universitäts-Verlag, Wiesbaden, 2000.
- [24] E. Feess and S. Ossig. Reimbursement schemes for hospitals, malpractice liability, and intrinsic motivation. *International Review of Law and Economics*, (27):423 – 441, 2007.
- [25] W. Fiori. Abrechnungsbetrug der Krankenäuser - Tatsache oder Fiktion. *Das Krankenhaus*, (1):17 – 32, 2010.
- [26] G. C. Fonarow und E. D. Peterson. Heart failure performance measures and outcomes - real or illusory gains. *JAMA*, 302(7):792 – 794, 2009.
- [27] G. D. Fraja. Contracts for health care and asymmetric information. *Journal of Health Economics*, (19):663–677, 2000.
- [28] E. Frese et al. „Diagnosis Related Groups“ (DRG) und kosteneffiziente Steuerungssysteme im Krankenhaus. *zfbf*, (56):737–759, 2004.

- [29] Bundesamt für Gesundheit BAG (Schweiz). Statistiken zur Krankenversicherung - Qualitätsindikatoren der Schweizer Akutspitäler 2007.
- [30] S. W. Glickman et al. Promoting quality: the health-care organization from a management perspective. *International Journal for Quality in Health Care*, 19(6):341 – 348, 2007.
- [31] S. W. Glickman, F.-S. Ou, and E. R. DeLong. Pay for performance, quality of care, and outcomes in acute myocardial infarction. *JAMA*, (21):2373 – 2380, 2007.
- [32] G. Gowrisankaran. *Competition, Information, and Hospital Quality*, Kapitel 12, 319–352. MIT Press, 2008.
- [33] H. Gravelle und R. Ress. *Microeconomics*. Prentice Hall: New York, 2004.
- [34] P. J. Hammond. Altruism. In: J. Eatwell, *The new Palgrave: a dictionary of economics*, Seiten 85 – 87. Macmillan, London, 1987.
- [35] B. Holmstrom und P. Milgrom. Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, & Organization*, 7; Special Issue:24–52, 1991.
- [36] B. Holmstrom und J. Tirole. The theory of the firm. In: R. Schmalensee and R. Willig, *Handbook of Industrial Organization*, Volume I, Kapitel 2, Seiten 62 – 133. Elsier Science Publishers, 1989.
- [37] M. Hoy et al. *Mathematics for economics*. MIT Press, Cambridge, Mass., 2. ed edition, 2001.
- [38] W. Jack. Purchasing health care services from providers with unknown altruism. *Journal of Health Economics*, (24):73–93, 2005.
- [39] I. Jelovac und I. Macho-Stadler. Comparing organizational structures in health services. *Journal of Economic Behavior & Organization*, 49:501–522, 2002.
- [40] S. F. Jencks et al. Quality of medical care delivered to medicare beneficiaries: A profile at state and national levels. *JAMA*, 284(13):1670 – 1676, 2000.
- [41] T. Kersting. Prozess und Struktur der Diagnostik und Therapie. In: B. Schmidt-Rettig and S. Eichhorn *Krankenhaus-Managementlehre: Theorie und Praxis eines integrierten Konzepts*. Kohlhammer, Stuttgart, 1. Aufl., 2008.
- [42] H.-F. Kienzle. Struktur- und Organisationsveränderungen in der Klinik aus Chefärztlicher Sicht. *ZaeFQ*, (101):527 – 530, 2007.

- [43] M. Kuhn und L. Siciliani. Upcoding and optimal auditing in health care (or the economics of DRG creep). *CEPR Discussion Paper No. 6689*, February 2008.
- [44] J.-J. Laffont und D. Martimort. *The Theory of Incentives*. Princeton University Press, 2002.
- [45] J.-J. Laffont und J. Tirole. *A theory of incentives in procurement and regulation*. MIT Press, 5 edition, 1993.
- [46] M. N. Lutfiyya et al. A comparison of quality of care indicators in urban acute care hospitals and rural critical access hospitals in the united states. *International Journal for Quality in Health Care*, 19(3):141 – 149, 2007.
- [47] C.-T. A. Ma. Health care payment systems: Cost and quality incentives. *Journal of Economics & Management Strategy*, 3(1):93–111, 1994.
- [48] A. Mas-Colell, M. D. Whinston und J. R. Green. *Microeconomic Theory*. Oxford University Press, 1995.
- [49] R. P. McAfee und J. M. Millan. Organizational diseconomies of scale. *Journal of Economics & Management Strategy*, 4(3):399–426, 1995.
- [50] E. A. McGlynn et al. The quality of health care delivered to adults in the united states. *The New England Journal of Medicine*, (348):2635 – 2645, 2003.
- [51] T. G. McGuire. Physician agency. In: A. Culyer und J. Newhouse, *Handbook of Health Economics*, Volume 1, Kapitel 9. Elsevier, 2000.
- [52] N. L. McKay und M. E. Deily. Cost inefficiency and hospital health outcomes. *Health Economics*, 17:833 – 848, 2008.
- [53] A. Mehrotra et al. Pay for performance in the hospital setting: What is the state of the evidence? *American Journal of Medical Quality*, 24(1):19 – 28, 2009.
- [54] D. Mookherjee. Decentralization, hierarchies and incentives: A mechanism design perspective. *Journal of Economic Literature*, 154:367–390, June 2006.
- [55] M. Mougeot und F. Naegelen. Adverse selection, moral hazard, and outlier payment policy. *Journal of Risk and Insurance*, 76(1):177–195, 2009.
- [56] A. Muthoo. *Bargaining theory with applications*. Cambridge University Press, Cambridge, 1999.
- [57] J. P. Newhouse. Toward a theory of nonprofit institutions: An economic model of a hospital. *The American Economic Review*, 60(1):64–74, 1970.

- [58] J. Palm. Prozessoptimierung durch klinische Pfade. *Führen und Verwalten im Krankenhaus (f&w)*, 26(5):493 – 496, 2009.
- [59] A. Radl. Digitales OP-Management im Landeskrankenhaus Bludenz. *Das Krankenhaus*, (6), 2009.
- [60] A. Rosado. Tinbergen’s rule. In: J. Segura and C. R. Braun, *An Eponymous Dictionary of Economics - A Guide to Laws and Theorems Named after Economists*, Seite 257. Edward Elgar Publishing Limited, 2004.
- [61] A. Ryan. Hospital-based pay-for-performance in the united states. *Health Economics*, (18):1109 – 1113, 2009.
- [62] B. Schmidt-Rettig. Leitungsstrukturen. In: B. Schmidt-Rettig und S. Eichhorn, *Krankenhaus-Managementlehre: Theorie und Praxis eines integrierten Konzepts*. Kohlhammer, Stuttgart, 1. Aufl., 2008.
- [63] J. Schreyögg, T. Stargardt, O. Tiemann und R. Busse. Methods to determine reimbursement rates for diagnosis related groups (DRG): A comparison of nine european countries. *Health Care Management Science*, (9):215 – 223, 2006.
- [64] Schwappach und Schubert. Offenlegen oder nicht? *Deutsche Medizinische Wochenschrift*, (132):2637–2642, 2007.
- [65] A. Shleifer. A theory of yardstick competition. *The Rand Journal of Economics*, 16(3):319–327, 1985.
- [66] L. Siciliani. Selection of treatment under prospective payment systems in the hospital sector. *Journal of Health Economics*, (25):479–499, 2006.
- [67] E. Silverman und J. Skinner. Medicare upcoding and hospital ownership. *Journal of Health Economics*, (23):369–389, 2004.
- [68] F. A. Sloan. Not-for-profit ownership and hospital behaviour. In: *Handbook of Health Economics*, Volume 1, Kapitel 21, Seiten 1141–1173. 2000.
- [69] A. Sonnentag. Trägerstrukturen und Rechtsformen. In: B. Schmidt-Rettig und S. Eichhorn, *Krankenhaus-Managementlehre: Theorie und Praxis eines integrierten Konzepts*. Kohlhammer, Stuttgart, 1. Aufl., 2008.
- [70] L. Steinmann. *Konsistenzprobleme der Data Envelopment Analysis in der empirischen Forschung*. Dissertationsschrift, Wirtschaftswissenschaftliche Fakultät der Universität Zürich, 2003.

- [71] W. Strüwe. Spitalfinanzierung 2012 - Datenschutz ade? *Swiss Medical Informatics*, (71):1–7, 2011.
- [72] N. Undritz. Krankenhaus. In: G. Kocher and W. Oggier, *Gesundheitswesen Schweiz 2004 - 2006. Eine aktuelle Übersicht*, 130 – 143. Verlag Hans Huber, 2004.
- [73] Universitätsspital Zürich. *Qualitätsbericht 08*, 2009.
- [74] D. Zahnd. Qualitätsindikatoren für die Schweizer Akutspitäler - Präsentation im Auftrag des Bundesamtes für Gesundheit. Präsentation, 2008.
- [75] B. P. Zietz, U. Buch, H. Heißmeyer und M. Karaus. Analyse von Administrativen Daten für das strategische und ökonomische Management. *Das Krankenhaus*, (1), 2011.
- [76] P. Zweifel und F. Breyer. *Health Economics*. Oxford University Press New York, 1997.